

# A corporate response to AIDS

Private-sector employers who face up to the HIV epidemic in Africa must be encouraged, not harangued.

Efforts to combat the AIDS pandemic in Africa and elsewhere are essentially public activities, led by national governments. Unfortunately, in many areas of Africa where the pandemic is most serious, the governments are barely functional. In some regions, there is no real authority in existence to help implement the monitoring, prevention and treatment of AIDS. In others, the body best placed to tackle the problem of AIDS may be the oil company, brewer or mining company whose presence dominates the local economy.

In this issue, *Nature* examines efforts being initiated by the oil and gas industry in one of Africa's most politically febrile and strategically important regions — the Niger River delta — to build a community-based programme that will limit the spread of AIDS (see page 140). This project is difficult to implement and fraught with risk. Yet it is an important experiment and, if successful, it could provide one way of tackling an immensely exacting challenge.

The Ibani-se initiative on Bonny Island, at the mouth of the Niger River, is being supported primarily by Nigeria Liquefied Natural Gas (NLNG), a joint venture between the Nigerian government and the energy companies Shell, ENI and Total. NLNG's main plant employs some 17,000 people and, at a cost of US\$15 billion so far, is thought to be the largest single fixed investment in Africa.

The scheme is intended to engage every element of the island's community, from sex workers to truck drivers and shopkeepers, in building a programme that will study the pandemic in detail while undertaking educational and self-help programmes to limit its spread. As the programme develops, it will make antiretroviral drug treatment available, free of charge, to those who need it.

One of the many tragedies of AIDS in Africa is its propensity to strike at the most economically active members of the population. Major hubs of activity such as the NLNG plant are particularly vulnerable to the ravages of the epidemic. Yet given the size of the population just outside the plant — tens of thousands of migrants have been drawn to Bonny Island alone — it is unrealistic to expect the plant's operators to support AIDS prevention and treatment programmes for the general population.

The best outcome, instead, is for companies such as NLNG to use their extensive logistical abilities to help forge a community-based

initiative that will eventually attract sustainable funding from outside sources, such as the Global Fund to Fight AIDS, Tuberculosis and Malaria, or the US President's Emergency Plan for AIDS Relief. That will require a project that is strongly identified with the private sector to forge an effective partnership with the public sector. This sounds simple but, as in many parts of Africa, it is in fact immensely difficult. Government officials at the local and national level have their own priorities and tend to assume that projects backed by corporations can look after themselves. The corporations, meanwhile, have learned to expect little in the way of competence or probity from public officials.

The multinational corporations have plenty to lose by squaring up publicly to the problem. Talk of the need for transparency and openness is cheap, but what employer really wants the painful facts about the extent of an AIDS epidemic among its employees to enter the scientific literature, or anywhere else in the public domain? At long last, major employers, such as multinational mining conglomerate Anglo American and NLNG, are biting the bullet in this regard and should be commended for doing so.

Private-public collaboration in tackling AIDS remains at an early stage of development, however. Some major companies still decline to participate, and the national committees that coordinate Global Fund programmes often remain unresponsive. The Global Fund won't tell these committees what to do — but it should prod them firmly in the direction of partnership with the private sector. Corporations, which are often the most viable institutions in their vicinities, have to accept some level of responsibility for what goes on there. But the penny has to drop that these companies, profitable as they may be, are not going fund health provision beyond that for their employees and immediate families.

In the West there is a widespread tendency to assume that the activities of multinationals in Africa are entirely malign. Some observers of political disquiet in Nigeria seem to think it would be better if Shell, for example, just packed its bags and went home. That is simply ridiculous. With AIDS, as in other spheres of social policy, the multinationals are not the problem. They should instead be seen as part of the solution. ■

## Off the rails

Aviation's role in climate change is causing a storm.

When a politician is accused of being "foolish", "ill-informed" and not having "a clue what he's talking about", you might expect the complaints to be legitimate. But such venomous language could also betray a different explanation: the politician may have rattled someone's cage. So it was last week, when aviation boss

Michael O'Leary reacted furiously to comments made by Ian Pearson, the British government's hitherto low-key environment minister.

Pearson had accused Ryanair, the Irish budget airline that O'Leary runs, of doing too little to tackle climate change, memorably branding the airline "the irresponsible face of capitalism". The fact that O'Leary protested rather too much was down to more than his well-earned reputation as a self-publicist. As pressure grows on the aviation industry, Ryanair and other airlines find themselves in an increasingly uncomfortable position.

The problem lies with two diverging industry trends. On the one

hand, flying is more popular than ever before. British passenger numbers are predicted to double to 475 million per year by 2030, for example, and in China, according to the World Bank, passenger numbers grew by a whopping 28% from 2003 to 2004.

Progress in improving the fuel efficiency of aircraft is, by contrast, gradual. Current trends suggest that gains of 1–2% per year will be the norm for the foreseeable future. Only a step change in aeroengine design might increase this, but with current approaches being so safe and reliable, and change so expensive, that seems improbable.

As a result, the growth in aviation emissions, if left unchecked, is liable to wreck attempts to reduce greenhouse-gas emissions. In Britain, for example, airlines' emissions are growing by around 7% each year, even as the government prepares to set in law a commitment to cut national emissions to 60% below 1990 levels by 2050. According to the Tyndall Centre for Climate Change Research, the rest of the economy would have to move to zero emissions in order to sustain aviation growth and meet the target.

It is clear that something will have to give. If real cuts are to be made in carbon emissions, aviation growth will certainly have to level off. The number of trips may even need to be cut. These are truths that the airline industry does not want to hear.

The first moves towards such a future are now being made. The European Commission announced in December that it plans to

include domestic flights in its carbon emissions trading scheme by 2011, with intercontinental flights joining the scheme the following year. This is a critical first step towards a future in which consumers start to pay for the environmental cost of flying.

For now, several European airlines say they are in favour of emissions trading, perhaps because they anticipate being eased gently into the scheme with relatively generous emissions targets. The full test of their support will come in the years after the scheme starts, when emissions will need to be capped tightly enough to reduce the growth in airline traffic.

But some, led by O'Leary's Ryanair, are opposed from the start. US airlines are sending an equally indignant message behind the scenes, Pearson says. This opposition may lead to a legal challenge from US airlines to the proposed inclusion of intercontinental flights.

European Commission officials say they are confident of the legality of their approach. But if the US legal action or non-cooperation of the airlines make emissions trading unworkable, more radical alternatives may have to be considered. One such approach would be to review European adherence to the 1947 Chicago Convention, the international agreement that prohibits the taxation of aviation fuel, and hence gives the industry a permanent advantage over its competition, such as rail travel. That would really give O'Leary and his allies something to squeal about. ■

## Competitive stumbling

Promised investment in the physical sciences is held up in a US budget jam.

In his State of the Union Address last January, President George W. Bush announced an American Competitiveness Initiative, which included substantial increases in funding for the physical sciences as a means of securing US industrial competitiveness in the long term. This laudable initiative set out to increase expenditure on research in physics, engineering and other disciplines at the National Science Foundation, the Department of Energy and the National Institute of Standards and Technology.

Leading Republicans and Democrats have expressed strong support for such a step, and as the president's proposal moved through the congressional budget process last year, all three agencies were looking forward to significant growth during 2007.

But despite the goodwill, an unusual turn in the budget saga has caused the gains to vanish overnight. The outgoing Republican Congress never finished the 2007 budget, as it should have done by 1 October 2006. That left the incoming, Democrat-controlled one to decide what to do with the unfinished appropriations bills. But the Democrats want to focus on the issues that they think got them elected, such as lobbying reform and the war in Iraq. Instead of completing last year's spending bills, the Democrat chairs of the appropriations committees have said they plan to stick with 2006 funding levels through the whole of the 2007 fiscal year (see page 130).

The announcements left lab directors and agency administrators who stood to benefit from the competitiveness initiative protesting

about sharp cutbacks in extramural grants and sidelined intramural facilities. Some of this is exaggerated, but there is no denying that the physical sciences will suffer as a result of this turn of events. Because the House and Senate had already approved the increases, agencies were incorporating them into budget planning. Tough choices must now be made at the last minute to ensure that the agencies can operate until 30 September, when the fiscal year ends. In other words, the sudden retraction of the increases is more damaging than if they had never been mooted in the first place.

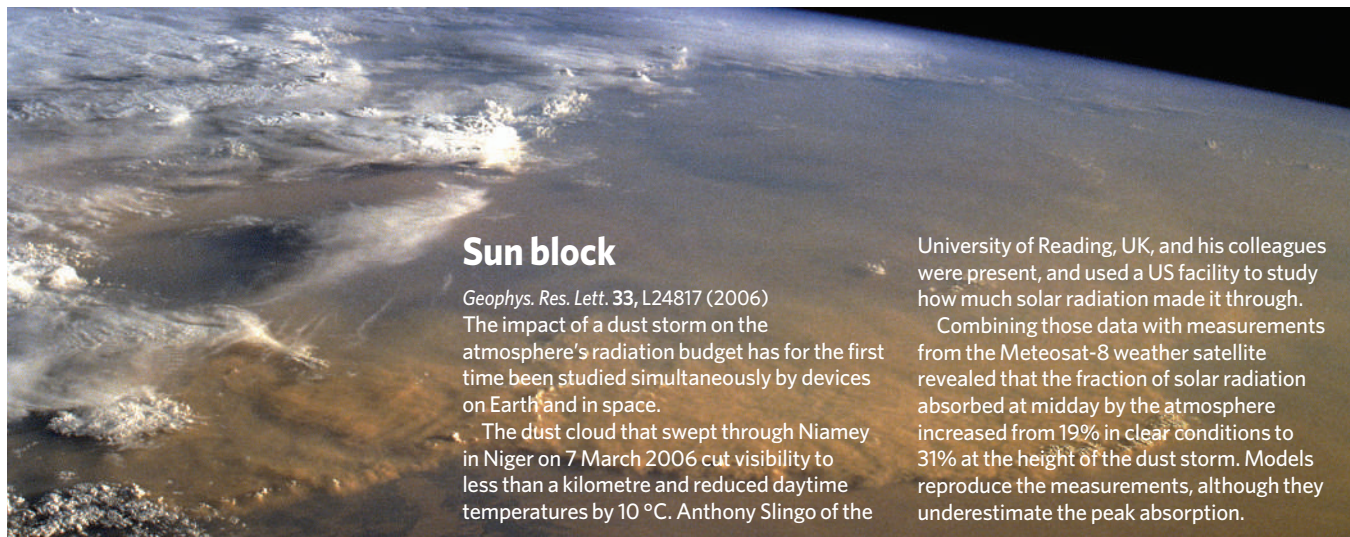
The bill to keep spending at 2006 levels will probably be finalized and passed in the next few weeks, and science lobbyists in Washington are scrambling to have an exception made for the competitiveness initiative. They have a strong case, and given the bipartisan support for the initiative's main elements, they deserve to succeed. Unfortunately, it is more likely that appropriators will make exceptions only for a few politically sensitive agencies, such as the Department of Veterans' Affairs.

All eyes will then turn to the president's budget proposal for 2008, to be released on 5 February. Despite the president's well-known political difficulties, his proposal will set the tone for the subsequent budget process. Science advocates have been pressing hard to ensure that Bush's proposal makes up for this year's shortfalls. And scientists should urge their own congressional representatives to ensure that the 2008 budget reflects both the withdrawn increase for 2007 and that originally envisaged for 2008. With the right support, it may still be possible to get the American Competitiveness Initiative back on track. ■

**"The sudden retraction of the funding increases is more damaging than if they had never been mooted in the first place."**



## RESEARCH HIGHLIGHTS



NASA

## Sun block

*Geophys. Res. Lett.* **33**, L24817 (2006)

The impact of a dust storm on the atmosphere's radiation budget has for the first time been studied simultaneously by devices on Earth and in space.

The dust cloud that swept through Niamey in Niger on 7 March 2006 cut visibility to less than a kilometre and reduced daytime temperatures by 10 °C. Anthony Slingo of the

University of Reading, UK, and his colleagues were present, and used a US facility to study how much solar radiation made it through.

Combining those data with measurements from the Meteosat-8 weather satellite revealed that the fraction of solar radiation absorbed at midday by the atmosphere increased from 19% in clear conditions to 31% at the height of the dust storm. Models reproduce the measurements, although they underestimate the peak absorption.

## PHYSIOLOGY

### Mighty mouse

*Cell Metabolism* **5**, 35–46 (2007)

A genetic tweak has converted mice into endurance runners by enriching a poorly understood form of muscle fibre. The discovery hints that 'type IIX' muscle fibres are an underappreciated contributor to athletic ability.

Bruce Spiegelman of Harvard Medical School in Boston, Massachusetts, and his team found that turning on a gene called PGC-1 $\beta$  in mice muscles transforms muscles that normally contain 15–20% type IIX fibres into ones with nearly 100%. This type of muscle fibre seems to have properties in between those of the better-known 'slow-twitch' and 'fast-twitch' varieties. The modified mice were able to run on a treadmill for 25% longer than normal before becoming exhausted.

The ability to modify muscle type could one day be useful in the treatment of muscle-wasting diseases.

derived stem cells did not form cancerous masses when left to grow on their own.

But the cells may not be able to differentiate into every cell type of the body and they share some characteristics with adult stem cells. This suggests that they might occupy a middle ground between embryonic and adult stem cells: not too flexible, nor too rigid, but just plastic enough to be useful in the clinic.

## PALAEONTOLOGY

### Turning over an old leaf

*Proc. Natl Acad. Sci. USA* **104**, 565–569 (2007)

The first example of a 'leaf insect' from the fossil record is presented by Sonja Wedmann of the University of Bonn in Germany and her colleagues.

Recovered from 47-million-year-old sediments, the fossil insect is a remarkably well-preserved male (pictured below). In size and overall shape it resembles modern leaf insects of the Phasmatodea order, which includes stick insects. That it shares some

characteristics with today's stick insects supports the idea that leaf-mimickers descended from stick-insect-like ancestors.

The new species' name, *Eophyllium messeleensis*, derives from the era to which the fossil dates, the Eocene, and the place where it was found in Germany, known as the Messel Formation.

## CANCER BIOLOGY

### The X factor

*Science* doi:10.1126/science.1137509 (2007)

Genes on the X chromosome may play an unappreciated role in cancer, say researchers in the United States.

Daniel Haber of the Massachusetts General Hospital in Boston and his colleagues studied tumours from patients with a kidney cancer called Wilms' tumour. They identified a new gene, *WTX*, on the X chromosome, which was missing or faulty in a third of the tumours tested.

*WTX* is a 'tumour suppressor' gene, which helps to stop cells becoming cancerous. Until now, such genes have been found only on chromosomes that cells carry in pairs, meaning that both chromosomes' genes must be faulty to trigger cancer. But only one X chromosome is ever active in normal cells. This means that only one mutation is needed to inactivate a tumour suppressor on the X chromosome.

## STEM CELLS

### Degrees of flexibility

*Nature Biotech.* doi:10.1038/nbt1274 (2007)

Stem cells derived from amniotic fluid could have just the right properties for use in future therapies, says a group led by Anthony Atala at Wake Forest University in Winston-Salem, North Carolina.

The researchers pulled cells with a stem-cell marker out of amniotic fluid left over from fetal diagnostic procedures. These stem cells could differentiate into many lineages — including bone and brain cells that functioned in live mice. And unlike stem cells derived from human embryos, the amniotic-fluid-



©2007 Nature Publishing Group

## ORGANIC CHEMISTRY

### An easy solution

*J. Am. Chem. Soc.* doi:10.1021/ja067487h (2007)

Chemists in the United States and Japan report the synthesis of two large, bowl-shaped organic molecules.

G. OLESCHINSKI, UNIV. BONN

The molecules are the polyarenes pentaindenocorannulene ( $C_{50}H_{20}$ ) and tetraindenocorannulene ( $C_{44}H_{18}$ ). Each consists of a sheet of carbon atoms arranged in hexagons, interspersed with pentagonal units that distort the sheet. The curvature around the central unit is greater than that seen in spherical carbon buckyballs, despite the presence of delocalized electrons, which typically act to flatten a molecule's structure.

Both molecules were made in solution. This suggests that nanotubes and buckyballs could be synthesized in solution using similar reaction steps, says the team led by Lawrence Scott of Boston College, Massachusetts. Currently, these structures have to be made by high-temperature gas-phase methods.

## IMMUNOLOGY

### Friend or foe?

*Nature Immunol.* doi:10.1038/ni1427 (2006)

Immune cells that can't distinguish friend from foe are the targets of a newly discovered immune-control mechanism operating in the gut.

Most self-reactive immune cells are weeded out before they mature, but a few escape that filter. Shannon Turley of the Dana Farber Cancer Institute in Boston, Massachusetts, and her colleagues describe a population of lymph node cells that identify these rogue immune cells by expressing self-antigens characteristic of intestinal tissue. Immune cells that react with the antigens are inactivated.

The other known mechanism for weeding out such escapees involves specialized cells that harvest self-antigens from neighbouring tissues. The researchers suggest that the new system, which provides an additional barrier to autoimmune attack, may also operate for other tissues.

## ASTRONOMY

### Stellar heavyweights

*Astrophys. J.* **654**, L143–L146 (2007)

The spectacular nebula NGC 6334 (pictured below) is studded with massive stars which, in some regions, are still forming. Luis Rodríguez of the Centre for Radioastronomy and Astrophysics in Morelia, Mexico, and his team may have caught a glimpse of this process.

The researchers used the Very Large Array telescope to image a region of the nebula known as NGC 6334 I(N). Their millimetre-wavelength pictures revealed four clumps of



dust many times wider than our own Solar System, with masses a few times that of our Sun. The team interprets these as disks from which stars are forming. The star or stars at the disks' centres are too obscured to see, but theory suggests they will grow by accretion into great monsters.

## GENETICS

### Spot the difference

*Nature Genet.* doi:10.1038/ng1955 (2007)

The genetic underpinnings of traits that differ between ethnic groups are giving themselves up to modern analysis techniques, a new

study demonstrates.

A group led by Vivian Cheung and Richard Spielman of the University of Pennsylvania, Philadelphia, compared cells from caucasian individuals with those from Japanese and Chinese people. Nearly 25% of the more than 4,000 genes they analysed were expressed at different levels in the two groups. The team then combed through the HapMap, which documents single-nucleotide polymorphisms, or SNPs, across various ethnic groups and identified genetic variations that seemed to account for the differences. The group concludes that genetically controlled variation in gene regulation is responsible for the phenotypic differences in the cell type they studied.

## PHYSICS

### No nanobubbles

*Proc. Natl Acad. Sci. USA* **103**, 18401–18404 (2006) and *Phys. Rev. Lett.* **97**, 266101 (2006)

Two independent teams say they have resolved a controversy about how water interacts with hydrophobic surfaces.

Past experiments have presented conflicting results: some showing

that a low-density layer of water forms where water meets the hydrophobic surface; others finding no such layer; and a handful suggesting that dissolved gases emerge to form 'nanobubbles' at the interface.

Now two teams have used X-rays to examine the junction between water and a layer of octadecylsilane in detail. One team collected data at the Argonne National Laboratory, Illinois, the other at the European Synchrotron Radiation Facility in Grenoble, France. Each team observed a 'depletion layer' of reduced density at the interface, which is in line with theoretical predictions; neither saw 'nanobubbles'.

## JOURNAL CLUB

**John Cowan**  
University of Oklahoma,  
Norman, Oklahoma, USA

**An astronomer invites you to contemplate the history of some of the oldest stars in the Universe.**

Much of my work is an attempt to determine what kinds of stars formed and what types of element synthesis occurred when our Galaxy was very young. This means that I am particularly

interested in a class of stars referred to as 'carbon-enhanced metal-poor'.

The composition of a star reflects the properties of the interstellar medium at the time it formed, which evolves as generations of stars come and go. Metal-poor stars were born early in the history of our Galaxy, before dying stars enriched the interstellar medium with heavy elements. The fact that some of these metal-poor stars are carbon-enhanced provides insight into the types of stars

that came before them.

Recently, one group reported that around 20% of metal-poor stars are carbon-enhanced (S. Lucatello *et al. Astrophys. J.* **652**, L37–L40; 2006). A previous study had produced a lower figure (J. Cohen *et al. Astrophys. J.* **633**, L109–L112; 2005), prompting a battle between the competing groups. But both papers agree that more metal-poor stars are carbon-enhanced than are younger, high-metallicity stars.

To me, this is one of the most interesting and compelling results

to come from the study of such stars. Massive stars — with at least ten times the mass of the Sun — produce carbon efficiently, so this gives us a clear indication that massive stars, although very rare today, were much more common early in the history of the Universe.

The differences between the studies' numbers may lie in how the authors define a carbon-enhanced star, or could be a matter of statistics. I look forward to future papers that address these issues — and perhaps continuing the controversy.



## NEWS

# Power shift stymies US science budget

## WASHINGTON DC

The new Democrat-led US Congress convened on 4 January, but the changeover in power from a Republican-dominated Congress has thwarted almost anyone looking for government money. A budget impasse has erased the gains expected by several research agencies, and science advocates have launched an all-out lobbying campaign to win back the increases.

At stake are large boosts proposed for the National Science Foundation (NSF), the Department of Energy's Office of Science and the National Institute of Standards and Technology (NIST). President George W. Bush proposed the increases last February, and Congress was supposed to hammer out the final numbers by 1 October 2006, the start of the 2007 fiscal year. But in December it became clear that the Republican Congress would be unable to complete the budget, leaving Democrats with the unsavoury task of finishing the Republican bills.

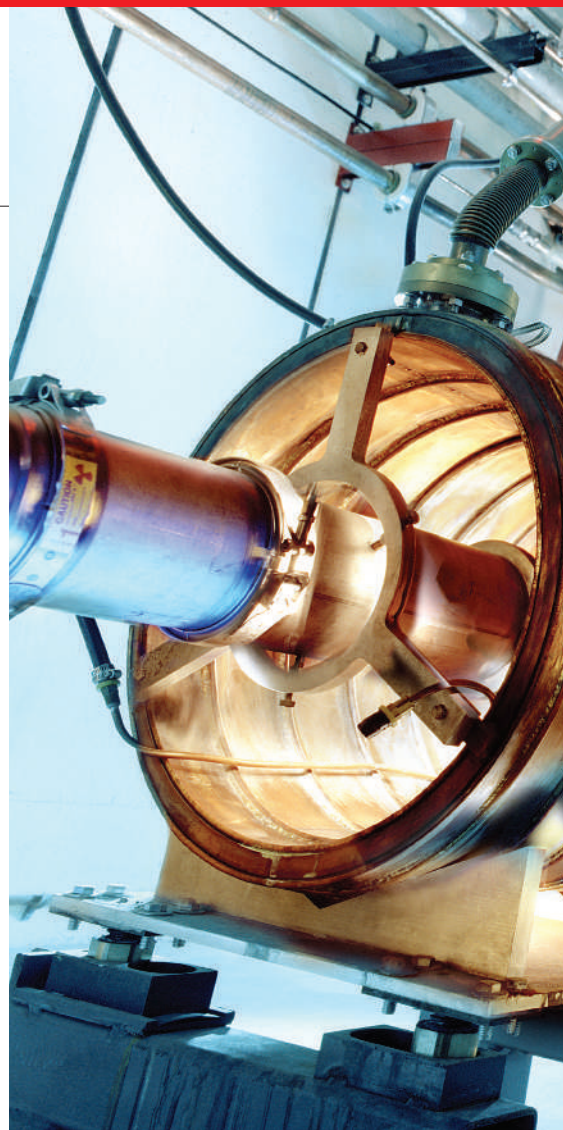
The new chairs of the committees that dole out the money said last month that they would abandon the budget in favour of a year-long "continuing resolution", which would keep funding for agencies at the previous year's levels. Now budget officials at science agencies are

**"Science is in a very precarious position. I think a lot of hopes and expectations will have to be put off."**

scrambling for options, which include cutting grants, delaying projects and even temporarily laying off staff. "Science is in a very precarious position," says Joel Widder, a science lobbyist with the firm Lewis-Burke Associates in Washington DC. "I think a lot of hopes and expectations will have to be put off."

Particularly affected would be agencies favoured by President Bush's American Competitiveness Initiative, which aims to bolster US innovation (see *Nature* **439**, 644–645; 2006). That includes the energy department's science office, which had been expecting a 15–18% boost in its \$3.6-billion budget. Meanwhile, Brookhaven National Laboratory in New York may

temporarily shut down its Relativistic Heavy Ion Collider, which was kept open last year only with the aid of a \$13-million private donation. At the Fermi National Accelerator Laboratory in Illinois, there are plans to shut down the Tevatron, the world's most powerful accelerator, and to lay off all staff without pay for a month, according to director Pier Oddone. And at the Stanford Linear Accelerator Center in California, construction of the Linac Coherent Light Source will be deferred for a year, says Keith Hodgson, the lab's acting director.



"You sound like you're whining, but the fact is that this comes after year after year of flat budgets," Hodgson notes. The science office hasn't seen a funding increase in roughly a decade.

The NSF would also lose up to an expected

P. GINTER

## Biomedical agency feels the pinch — again

For the US National Institutes of Health (NIH), the budget problems in the US Congress could not have come at a worse time. The agency is already struggling to cope with a budget that shrank nearly 7% in real terms between 2004 and 2006, even as the number of scientists applying for NIH funds soared (see *Nature* **443**, 894–895; 2006).

"This is not a one-time hit we're taking," says John Niederhuber, director of the National Cancer Institute, the agency's largest organization. "The budget is under as much stress as I have experienced in my 30-plus-year career as an extramural investigator."

The House of Representatives

was looking to cut the NIH budget by 0.3% from last year, whereas the Senate would have raised it by 0.7%. The impasse in Congress means that the budget is actually likely to be frozen at its 2006 level of \$28.6 billion until the end of September. That figure does not factor in biomedical inflation, which the Department of Commerce gauged at 3.5% in 2006. "In effect, we are cutting NIH purchasing power by another 3.5%," says Jon Retzlaff, director of legislative relations at the Federation of American Societies for Experimental Biology (FASEB).

The cuts are being felt on the ground. In October, extramural investigators were notified that

routine grant renewals would be paid at 80% of promised levels. "The people who suffer the most pain are the junior staff," says John Moore, an AIDS researcher at Weill Cornell Medical College in New York City, who notes that as a senior investigator he has other grants and reserves to call on.

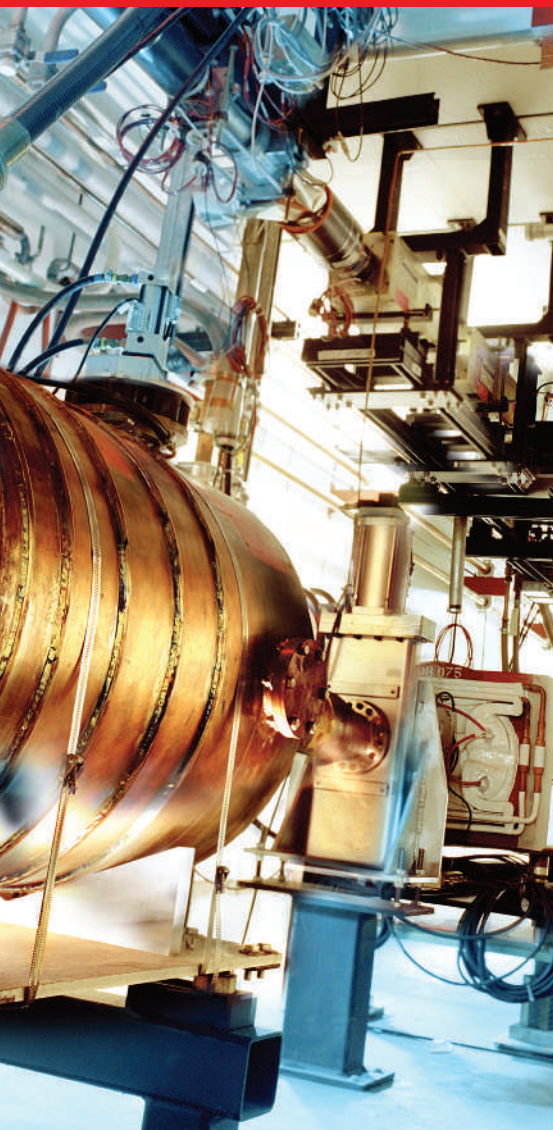
Sally Rockey, deputy director of the office of extramural research at the NIH, points out that the current funding at 80% is a conservative strategy adopted until 2007 funding is finalized. "Should we get a budget, we would consider upward adjustments of those levels," she says. "And we have every reason to expect that we will have an appropriation this year."

Once a budget is finalized, she says, investigators should expect funding for routine grant renewals to be closer to the roughly 97% level in place before October.

NIH lobbyists are hoping that the agency will get some of the billions that may be freed up by a congressional plan to eliminate funding for lawmakers' pet projects in 2007. "We are cautiously optimistic," says Leo Furcht, the president of FASEB. Retzlaff says that between \$500 million and \$1 billion is the most that the NIH could hope to see from this.

Niederhuber is more circumspect. "There are a lot of tin cups out there," he says.

**Meredith Wadman**



**Power cut?** The Tevatron particle accelerator in Illinois, the world's largest, may be suspended.

7.8% boost in its \$5.6-billion budget. The agency was hoping to use the funds to increase the number of grants and launch its participation in the International Polar Year, says Jeff Nesbit, the NSF's director of legislative and public affairs. Under flat funding, he estimates that the agency would eliminate around 600 new grants, 10% of the number it was hoping to grant next year. Polar-year activities would also be jeopardized (see 'Shrunken heavyweight undermines polar year'), as would the start of a new petascale computing centre. Plans for the new EarthScope programme of geological monitoring might have to be slowed, and travel has been severely curtailed.

At NIST, a planned internal reorganization is now in question. The \$80-million Advanced Technology Program, which funds outside research, had been expected to be cut, with that amount of money and more being reallocated into a diverse portfolio of internal laboratories. The staff, though, are apparently not too disappointed. "Proposed budgets rarely come through as proposed, so there were no emotions here," says spokesman Michael Baum.

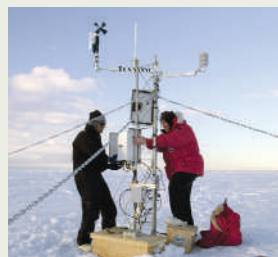
## Shrunken heavyweight undermines polar year

The International Polar Year (IPY), which starts on 1 March, is set to be the most comprehensive polar research programme for 50 years. But budget problems mean that the United States won't be the expected heavyweight for the two years of studies.

After three years of planning, the lead US agency — the National Science Foundation (NSF) — was to have provided some US\$62 million per year of new funds for the IPY. But only about \$15 million shifted from other NSF programmes may be available, officials say.

"It is tragic that the United States is not able to take a leading role," says atmospheric scientist Chris Rapley, director of the British Antarctic Survey and a lead planner for the IPY. "We always knew everyone might not be able to participate fully. But it's more damaging when the United States is not a major player."

There hasn't been such a coordinated polar research plan since the International Geophysical Year in 1957–58. The IPY involves more than 60 scientists from 35 nations, drawing data from research ships, satellites, ice stations



and marine outposts, to study climate change, for example. But a lack of full US funding will mean missed opportunities, says the University of Alberta's David Hik, a leading Canadian planner.

So far, about 12 nations have agreed to contribute \$200 million towards IPY studies. Canada is now the biggest contributor,

at US\$128 million. Other nations range from Norway (\$40 million), Russia, France and Britain (\$10 million each), to Brazil (\$4 million), Chile (\$2 million) and Malaysia (\$1 million), where there are concerns over the sea-level rise from melting ice.

Among the projects delayed is one led by David Bromwich, an atmospheric scientist at Ohio State University in Columbus. He had planned to fly the HIAPER research plane from the tip of South America to Antarctica to study processes related to global warming. That trip is now delayed beyond the two-year IPY window.

And Robin Bell, a geophysicist at Columbia University who chairs the National Academy of Sciences' polar research board, has seen her own project to explore Antarctic mountains delayed until the end of the IPY. "There is discouragement," she says. **Rex Dalton**

At NASA, small cuts in its \$16.8-billion budget could force further delays in the development and launch of several spacecraft, says Paul Hertz, who heads the agency's Science Mission Directorate.

### And the winners are...

The continuing resolution isn't bad news for everyone. The Department of Agriculture and the Environmental Protection Agency will both avoid proposed cuts. The US Geological Survey will avoid having to slash its minerals resource programme, which finds and describes materials such as copper, sand and gravel. And the National Oceanic and Atmospheric Administration had been set to be a bone of contention between the House of Representatives and the Senate. The House had wanted to cut the agency's budget by around \$500 million, whereas the Senate wanted to boost it by \$536 million. Now the agency will probably make do with its originally proposed level of \$3.7 billion.

Congress has managed to pass final 2007 budgets for only two departments — defence and homeland security. The Pentagon will see its budget for basic and applied research rise by 1.7% to \$6.8 billion, while the Department of

Homeland Security's budget for fundamental science would fall by 56% to \$42 million.

For other agencies, the proposed cutback to 2006 funding levels is not finalized (see 'Biomedical agency feels the pinch — again'). The new appropriations committees have yet to draft the resolution that would have to be signed into law, and they are considering exceptions, says Tom Gavin, a spokesman for Robert Byrd (Democrat, West Virginia), who heads the Senate appropriations committee. In the coming weeks, congressional subcommittees will consider providing agencies with additional funds. "There will be no set formula," Gavin says. "It will be on a case-by-case basis."

Science advocates are mounting an eleventh-hour campaign to have their case heard by the appropriators. "We're encouraging our universities to make contact with their individual members," says Tobin Smith, a federal-relations officer at the Association of American Universities. But the ultimate outcome remains far from clear, he notes: "I haven't talked to anyone who knows what's going to happen."

**Geoff Brumfiel, with additional reporting by Emma Marris**

See Editorial, page 123.

J. LANDIS, NSF



## SPECIAL REPORT

# Out of bounds

With the use of transgenic crops expanding around the globe, we need to decide what level of unapproved plants we are willing to accept in our diets. Zero is not an option, says **Heidi Ledford**.

Steve Linscombe still isn't quite sure how it happened. The director of the Louisiana State University AgCenter for Rice Research knows that he grew a few lines of transgenic rice in field trials between 2001 and 2003. He also knows that one of those lines, LLRICE601, was grown on less than one acre. What he is not clear on is how the line then wended its way into the food supply. That little mystery is now the subject of an official investigation and a class-action lawsuit.

When the escape was announced in August last year, LLRICE601 had not been approved for human consumption. The US Department of Agriculture (USDA) rushed to deregulate the crop, granting permission on 24 November for LLRICE601 to be grown without a permit. By then, Japan had already declared a month-long ban on all imports of US long-grain rice, and the European Union had started to require all US long-grain rice imports to be tested and certified at the expense of the exporters. Meanwhile, Bayer CropScience, the company that created the rice strain, put the blame squarely on farmers and an "act of God".

By that logic, this would not be the first time that a deity has aided and abetted the escape of a genetically engineered crop. On 21 December, Syngenta was fined \$1.5 million for allowing its unapproved pest-resistant Bt10 corn (maize) to mix into seed distributed for food. The past decade is smattered with examples of unapproved crops sneaking through containment barriers (see 'Some past escapes'). When they make it into the food supply — as with LLRICE601 and Bt10 — public outcry and financial losses follow. But amid the calls for tighter regulations, experts say one truth is being drowned out: no

amount of regulation can guarantee that these crops will not escape and multiply.

Meanwhile, the stakes are getting higher. Since 1991, the USDA has approved nearly 400 field tests of crops that produce pharmaceutical and industrial compounds, leaving many concerned that future escapes could have severe consequences for human health. A close call came in 2002, when stalks of corn designed to produce a pig vaccine were found mixed with \$2.7-million worth of Nebraska soya beans destined for human consumption. Prodigene, the corn's maker, was fined \$250,000 and forced to buy and destroy the soya beans.

## No guarantees

Although the use of transgenic crops is spreading around the globe, production is still concentrated in the United States, which grows more than half of the world's genetically engineered crops. There, they are monitored by three regulatory agencies: the USDA regulates field tests, the Environmental Protection Agency monitors crops genetically engineered to produce pesticides, and the Food and Drug Administration provides a voluntary 'consultation' on the safety of crops for human consumption. That voluntary consultation sets the United States apart from many other countries, including China and many European countries, which require crops to be evaluated for toxicity and allergenicity before being approved.

In the 20 years since the USDA started to regulate field tests, it has approved nearly 50,000 field sites. But an internal audit commissioned by the USDA inspector-general and released on 22 December 2005 was severely critical. The report admonished the agency for lacking basic

information about test sites, failing to inspect field tests sufficiently, and neglecting the fate of the crops after testing. USDA regulator Rebecca Besch says that a year on, many of the report's recommendations have been enacted. The agency now asks for detailed coordinates of field test sites, she says, and is revising its environmental standards.

Jeffrey Wolt, an agronomist at Iowa State University in Ames, commends the USDA for its efforts, but says that tougher regulations are no guarantee of confinement. "There has been this strong effort by regulators and industry to tighten this stuff up," he says. "But no matter how much you ratchet it down, the risk is not going to be absolute zero because that's a scientific impossibility."

Other scientists agree. Transgenic plants have many ways to escape. For plants pollinated by wind and insects, such as canola, pollen transfer is a constant threat. And although seed harvesting and processing equipment is designed to keep different varieties apart, there is no guarantee of success. "Just like anything, it is not 100%," says Linscombe. "You could have a seed that gets caught somewhere in a planter and later jars itself loose." And of course even if



## SOME PAST ESCAPES

Limagrain Seed and Monsanto withdrew 60,000 bags of Canadian canola after finding that it was contaminated with unapproved herbicide-resistant seed.

Unapproved insect-resistant corn produced by Monsanto escaped its field trial site and released pollen to a commercial crop. The commercial corn was destroyed.

ProdiGene field-tested corn in 2001 that was engineered to produce a pharmaceutical. The next year, transgenic corn was found mixed with commercial corn that surrounded the site, and the crops were destroyed.

Transgenic contaminants of corn engineered to produce a pharmaceutical were harvested with commercial soya bean plants a year after they were field-tested by ProdiGene. Some 500,000 bushels of soya beans were destroyed.

Unexpected winds at a field-test site released herbicide-resistant, transgenic bentgrass produced by Scotts Company beyond its containment area.

1997

2001

2002

2002

2004





**Escape route:** genetically modified corn can mix with other crops if not harvested properly.

only a few seeds make their way into breeding stock, their numbers can then multiply.

Meanwhile, says Michelle Marvier, an ecologist at Santa Clara University in California, the focus on designing effective biological containment has kept attention away from an even more slippery culprit: human error. "The reality is that humans are involved, and we inevitably make mistakes." She warns that any risk evaluation of a genetically engineered crop should consider that crop likely to escape.

Several countries have opted not to take that risk. After the news of LLRICE601 contamination, major exporters in Vietnam announced that they would not be growing any transgenic rice. And even some countries that grow genetically modified crops are cautious about the ones they will accept. Argentina, for example, the world's second largest producer, refuses to grow any genetically engineered crop that has not been approved for consumption in its major export markets, including the European Union. That policy is intended to prevent unintended mixing of crops from hurting Argentina's

robust agricultural export sector (although it hasn't protected neighbouring Brazil — which did not allow genetically engineered crops until last year — from repeated contamination from Argentina's transgenic stocks).

### Harsh punishments

In the United States, the idea is that escape can be prevented if producers know that they will be punished if unapproved plant material is detected in the food supply. If a company is responsible for contamination, it typically has to remove the unapproved material at its own expense, and as an additional deterrent, deal with the flurry of negative press that undoubtedly follows. "It is really bad for the reputation of these firms and the technology itself," says Guillaume Gruere, an agricultural economist at the International Food Policy Research Institute in Washington DC. But he says that the regulatory agencies' zero-tolerance policy clashes with the inevitability of escape. "The problem is the threshold. If you want zero percent, it's going to be pretty much impossible."

And despite the negative press, US public opinion of genetically modified crops seems to have been changed little by the escapes so far, judging from the results of a survey done by the Pew Initiative on Food and Biotechnology last year. The number of Americans who approve of genetically modified food has hovered unchanged at around 26% for the past five years, whereas the number that explicitly disapproves has shrunk from 58% to 46%.

What about crops that produce pharmaceuticals and industrial compounds? In 2003, the USDA issued stricter guidelines for containment of these plants. Isolation distances from food crops were increased, and field test sites were to be inspected more frequently. And so far, no such strains have been deregulated, meaning that they must always be contained no matter how well they are tested.

But Margaret Mellon, director of the Union of

Concerned Scientists' Food and Environment programme, doubts that those regulations will be enough. Her organization has called for a ban on the outdoor production of pharmaceuticals in food crops, arguing that the amount of regulation needed to guarantee containment would be prohibitively expensive both to the government and to researchers. The union has gone through the USDA regulations and analysed points at which transgenic crops could still escape, such as machine cleaning and seed transport. "Regulations that are sufficiently stringent to plug all of those holes really are not feasible," says Mellon. "We see how much trouble the agency is having even with the current ones." Instead, she argues, production of pharmaceuticals or industrial compounds should take place only in non-food crops such as tobacco.

The problem is that such a ban would have a chilling effect on research, because the technology for creating and processing transgenic food crops is well understood and therefore much cheaper. At this stage, a US ban seems unlikely, and no other country has an official ban on pharmaceutical-producing crops.

In 2005, the Oregon Department of Agriculture convened a panel to evaluate the risks and potential economic benefits of growing animals and plants that produce pharmaceuticals in the state. The panel concluded last October that the benefits outweigh the risks.

Back in Louisiana, Linscombe plans to enact a few new regulations of his own. After his experience with LLRICE601, he says that he will be taking drastic measures to separate any experimental crops from his breeding stock, to at least minimize the chance of contamination. He is considering buying separate processing equipment for genetically modified crops. And he plans to greatly exceed the typical three-metre distance that is required between strains. "We have two farms that are located five miles apart," he says. "Any transgenic work in the future is going to be on one farm, and the breeding work on the other."



### GM CROPS IN FOCUS

Find all our news coverage on transgenic crops in one place.

[www.nature.com/news/infocus/gmcrops.html](http://www.nature.com/news/infocus/gmcrops.html)

R. PEDRONCELLI/AP

T. W. EGGERS/CORBIS

Nature reported that Syngenta had mistakenly produced and distributed a regulated, insect-resistant strain of genetically modified corn (434, 423; 2005). The Environmental Protection Agency and the USDA decided that the crop did not pose a risk to human health.

Greenpeace reported that it had found evidence of unapproved rice being sold illegally in China over the previous two years.

The USDA found that BASF had planted regulated genetically engineered corn outside the area specified by its permit.

Unapproved herbicide-resistant rice produced by Bayer CropScience was found in US rice sold for food. The USDA decided retroactively that the crop did not pose a risk to human health.

Unapproved pest-resistant transgenic rice was found in European imports from China.

2005

2005

2006

2006

2006



# Congress and Bush set to clash on stem cells again

In a direct but possibly futile challenge to the only veto of George Bush's presidency, the US Congress is expected to pass a bill for a second time that would dramatically expand researchers' access to human embryonic stem-cell lines. As *Nature* went to press, a vote in the House of Representatives was expected on 11 January, with the Senate likely to act on an identical measure within a few weeks.

The bill would allow federal funds to be used for research on stem-cell lines derived from surplus embryos at fertility clinics and otherwise slated for destruction. The embryos would have to be donated with informed consent and without payment or other inducements.

The California-based Rand Corporation estimated in 2003 that about 11,000 of the roughly 400,000 embryos stored at US fertility clinics at that time were available for research, from which it should be possible to derive up to 275 cell lines (D. I. Hoffman *et al. Fertil. Steril.* **79**, 1063–1069; 2003). The study's lead author, David Hoffman of IVF Florida Reproductive Associates in Margate, estimates that there may now be 10–15% more embryos in storage. Under current Bush administration policy, only around 20 stem-cell lines, all created before 9 August 2001, are available to federally

funded researchers. These ageing lines are contaminated by having been grown on 'feeder' layers of mouse cells, making them impractical for developing human therapies.

Bush vetoed an identical bill last summer after both the House and Senate passed it, and the House failed by 51 votes to muster the two-thirds majority needed to override his veto (see *Nature* **442**, 335; 2006). Bush is likely to veto the bill again, and even with Democratic election gains in November, it is doubtful whether stem-cell supporters in the House have the votes to override Bush. Just 13 of the House members who voted with Bush last summer were defeated in November. And seven of the newly elected Democrats are anti-abortion.

But in the Senate, which passed the bill by 63 to 37 last summer, November's elections ushered in a net gain of three senators who are likely to approve of the measure, bringing supporters within one vote of a veto-proof majority. If the Senate did override a second Bush veto, pressure on House members to do the same could substantially increase.

The Democrats, newly in charge of the House and Senate, have made it an early priority to reintroduce the bill, now dubbed the Stem Cell Research Enhancement Act of 2007. "There is growing support for ethical embry-

**"We will use every legislative means at our disposal to ensure that the bill is enacted into law. It will happen during this Congress."**



Diana DeGette sees a public consensus in favour of federal funding of embryonic stem-cell research.

onic stem-cell research in Congress, which is caused by a solid public consensus supporting this research," Representative Diana DeGette (Democrat, Colorado), one of the bill's two House authors, told *Nature* shortly before she introduced the bill on 5 January.

Opponents such as Senator Sam Brown-

## Indian science is in decline, says prime minister

### BANGALORE

For several years, criticisms have been aired at India's scientific meetings and in its journals. Now the country's top politician has spoken out about the talent crisis and poor standards affecting Indian science.

Prime Minister Manmohan Singh voiced his feelings on 3 January, when he opened the 94th Indian Science Congress (ISC) in Chidambaram in southern India. The annual meeting attracts thousands

of scientists from across the country. Traditionally opened by the prime minister, it is generally an occasion to praise Indian science. But this year was different.

"While our government will do its utmost to invest in science, I call upon the scientific community to also invest its time and intellectual energy in the revitalization of our science institutions," said Singh. He added that he is "deeply concerned" about declining

enrolment in basic sciences, and said that the decaying university system "needs upgrading in a massive way".

Singh was most troubled, he said, by "the decline in the standards of our research work in universities and even in advanced research institutes". For example, the return on billions of rupees invested in alternative energy sources was inadequate. "Be it [hydropower], thermal or nuclear power, we have to improve the productivity

of investments already made." To raise quality, Singh warned that he may submit India's research labs to international peer review.

This is the first time that Singh has publicly criticized Indian science. His comments were apparently provoked by briefings from his science adviser, the distinguished chemist C. N. R. Rao, indicating that India publishes only 2.7% of the world's science papers (compared with China's 6%) and describing the flight of talent



D. ZALUBOWSKI/AP

back (Republican, Kansas) are not persuaded. "Right now, the private sector can spend all it wants on destructive human embryonic stem-cell research, but such research is producing no human applications. There is simply no need to waste taxpayer dollars on this," says Becky Ogilvie, a spokeswoman for Brownback.

Critics arguing that human embryonic stem-cell research is unnecessary were given ammunition on 7 January when *Nature*

*Biotechnology* published a paper from a group led by Anthony Atala, director of the Institute for Regenerative Medicine at Wake Forest University in Winston-Salem, North Carolina. The researchers reported isolating stem cells from human amniotic fluid that can differentiate *in vitro* into many of the same major cell types as embryonic stem cells (P. De Coppi *et al.* *Nature Biotechnol.* doi:10.1038/nbt1274; 2006).

The study provides "yet another reason why there is no need to destroy young human embryos in order to obtain their biological parts", Brownback said in a statement. Supporters of the bill shot back. "This study...in no way replaces the need for expanding the [US government's] embryonic stem-cell policy," Representative Michael Castle (Republican, Delaware), the other author of the House bill, said in his own statement.

Some supporters hope to amend the bill to make it more politically palatable to hesitant senators and more difficult for Bush to veto — perhaps by incorporating support for research on stem cells from adult sources.

But Bush has adamantly opposed expanded federal funding. DeGette notes that she asked to meet the president in November. Last week, she got a response from Bush's appointments secretary. It read: "The president would appreciate an opportunity to visit with you. Regrettably it will not be possible for us to arrange such a meeting at this time."

Democratic senator Tom Harkin of Iowa, a leading supporter of the bill, vows to fight a second veto. On 4 January, the day the bill was introduced in the Senate, he said: "We will use every legislative means at our disposal to ensure that [the bill] is enacted into law. And it will happen during this Congress."

**Meredith Wadman**

from government labs and universities to industry and abroad.

"In the past year our government has launched three new research institutions," he told his audience, adding that the government is considering creating more. He also vowed to increase annual expenditure on science and technology from less than 1% of gross domestic product to 2% over the next five years.

India's president A. P. J. Abdul Kalam welcomed Singh's assurance of extra science funding, but in his own address to the



Manmohan Singh: offered more money and institutes.

congress on 5 January pointed out that more than 20% of the country's science and technology budget in 2005–06 wasn't spent. He called for a plan for how future money will be used.

Rao, who heads the Jawaharlal Nehru Centre for Advanced Scientific Research in Bangalore, says attracting and keeping scientists is tough when companies offer up to six times the salary of

government labs: "In two years, I lost eight of my postdocs to the General Electric Company research centre next door."

**K. S. Jayaraman**

H. SAUKKOMAA/LEHTIKUVA/REUTERS



## ROBOT NEWS

### Noelle nativity

Hampered by South Korea's low birth rate, medical students in need of obstetric experience are turning to Noelle, a mannequin mother complete with robotic baby equipped with pink and blue lights to indicate its 'health'.

### Faking it

Game wardens in New Hampshire made more than 50 arrests for illegal hunting in 2006, with the aid of motorized decoy deer featuring realistic head and tail movements.

## NUMBER CRUNCH

**5** Atlantic hurricanes occurred in the 2006 season.

**79** was their combined ferocity on the Accumulated Cyclone Energy index.

**81** was the energy of the single biggest 2006 Pacific storm, Super Typhoon Ioke.

## ON THE RECORD

**“We are very excited about this game as it presents both sides of the creation-evolution argument, and so shows that the contemporary theory of evolution is perhaps the greatest hoax of modern times.”**

Erstwhile actor and born-again Christian Kirk Cameron, creator of the board game Intelligent Design vs Evolution, clearly doesn't trust players to make up their own minds.

**“Elephants around the country will enjoy a delicious lunch today consisting of about five Christmas trees each.”**

Berlin Zoo spokesman Ragnar Kühne on a festive recycling solution.

Sources: Reuters, VAOnews.com, Wunderblog, Journal Chrétien



M. RIETSCHEL/AP



Holders of foreign passports are finding their access to the West Bank denied or restricted by Israeli border controls.

M. AL-SHAER/AFP/GETTY

# Immigration policy forces researchers out of West Bank

A few kilometres outside the West Bank city of Ramallah, not far from the rubble-strewn compound where Yasser Arafat lived for many years, lies the campus of Birzeit University. As elsewhere on the West Bank, Birzeit's students and staff have learnt to live with the disruption and casualties caused by the occupying Israeli military and armed Palestinian groups. But now the university says that it faces a non-violent but potentially more dangerous threat to its existence: an Israeli policy that is forcing academics to abandon their jobs.

Around a year ago, say Palestinian academics, people with foreign passports started to find that their access to the West Bank and Gaza Strip was either denied or restricted by Israeli border controls. That poses a serious problem for universities. As there is no such thing as a Palestinian passport, researchers will, given the chance, often become naturalized citizens of other nations — forfeiting their right to Palestinian identification papers. Yet the new policy means that those academics, together with foreign staff, can find that a trip abroad ends in involuntary exile.

At Birzeit, officials say that they cannot put an exact figure on the number denied re-entry, but say the policy is one reason that around half of the university's 57 foreign academics had to leave last year. Campaigners say that other universities in the regions have suffered the same consequences, although precise data are not available. “We don't know why they are doing this, but brain drain will be the end product,” says Sarit Michaeli, communications director for B'Tselem, an Israeli human-rights group based in Jerusalem.

The rationale for the move is hard to confirm, as the Israeli government initially said

that it was simply enforcing existing rules. Many holders of foreign passports no longer have residence permits for the occupied territories and so have to repeatedly obtain three-month tourist visas to live there. Israeli officials said that the restrictions were designed to end this misuse of tourist visas. A request for clarification of this point received no reply.

But in meetings with foreign diplomats last month, Israeli officials seemed to acknowledge the problem, saying that they would start to renew tourist visas and allow access to those whose passports have been stamped “last entry”. Palestinian campaigners say they have since seen written confirmation of this intent, but note that they have seen no sign of a change in policy at the borders and that some foreign-passport holders are still being refused entry.

If the restrictions continue, academics from both sides of the border say they will fight them. A small group banded together last October to form the Israeli Committee for the Right of Residency, which is attempting to persuade Israel's academics to take a stand against the restrictions. Jacob Katriel, an emeritus physics professor at Technion, the Israeli Institute of Technology in Haifa, says the call was well received when he presented it last month during a session on human rights at the Statistical Mechanics Conference at Rutgers University, New Jersey.

Palestinian universities are also asking foreign academics and scientific organizations to protest against the rules. They say that the loss of academic talent, together with a drop in fees earned from foreign students, could destabilize the fragile education and research infrastructure of the West Bank and Gaza Strip. ■

Jim Giles

## Security lapses cost nuclear chief his job

The head of the US nuclear-weapons programme has been sacked.

Linton Brooks, head of the National Nuclear Security Administration since 2002, was fired on 4 January. Energy secretary Samuel Bodman cited security problems at Los Alamos National Laboratory in New Mexico as his main reason for dismissing Brooks. The lab has had several security lapses, most recently in October, when a computer drive containing classified information was found in a drug dealer's trailer.

Brooks expressed regret at the decision, but said that it was "based on the principle of accountability that should govern all public service". Taking Brooks's place temporarily will be Thomas D'Agostino, the agency's deputy administrator for defence programmes.

## Capitol Hill anthrax attack 'hit wide area'

The anthrax mailed in 2001 to the office of US senator Tom Daschle affected more people than previously thought, researchers have found.

A team led by biologists at the Naval Medical Research Center in Silver Spring, Maryland, measured immune responses in 124 people after the bioterrorist attack. Of those, 83 were near or in Daschle's office when the anthrax-laced letter was opened, and another 20 were farther away and had presumably not been exposed.

But even people standing outside the building when the spores were released generated immune responses against anthrax, suggesting that they had been exposed, the team found (D. L. Doolan *et al.* *J. Infect. Dis.* 195, 174–184; 2007).

The team also showed that cellular immune responses declined after treatment,

## Amazon founder's rocket lifts off

A secretive space tourism project became a little less so this week, after Amazon founder Jeff Bezos and his Blue Origin team released a video of a test launch done last November in Texas.

The grainy footage (see [www.blueorigin.com](http://www.blueorigin.com)) shows the prototype spacecraft (right), powered by hydrogen peroxide, blasting off vertically and returning to Earth the same way up, in true 1950s science-fiction style. The craft ascended about 85 metres — hardly outer space. But Bezos hopes that the design will eventually offer an efficient way to take well-heeled flyers beyond the atmosphere.



BLUE ORIGIN

suggesting that the antibiotics were effective against the bacterium — something that has never been shown for a bioterrorist attack.

## Biologist plans hunger strike over tenure snub

A biologist is threatening to go on hunger strike after being denied tenure at the Massachusetts Institute of Technology (MIT) in Cambridge, alleging that he has been discriminated against because of his race.

James Sherley, a black associate professor who works on stem cells in MIT's biological engineering division, intends to begin his protest on 5 February. He asked for colleagues' support in a letter dated 19 December. "My plea to the MIT faculty is not 'please, help me get tenure at MIT'. It was, and is, 'help me end racism at MIT!'," he says.

MIT has issued a statement saying that it is confident that its procedure for reviewing and granting tenure "was followed with integrity in this case". The decision has already been reviewed independently by senior faculty members.

## ExxonMobil accused over strategy on climate change

The Union of Concerned Scientists has issued a fresh set of charges against ExxonMobil's alleged campaign to spread doubt about global warming.

Drawing on the US oil company's reports and other documents, the group says that between 1998 and 2005, ExxonMobil gave US\$16 million to bodies dedicated in part to amplifying public perceptions of scientific uncertainty over climate change. The organizations include well-known groups such as the Competitive Enterprise Institute and smaller ones such as Frontiers of Freedom.

ExxonMobil says that many of the report's conclusions are "inaccurate".

## Venezuela legislates to boost research spending

Industrial companies in Venezuela will be compelled to invest in research and development, under a law that took effect on 1 January.

Companies that have an annual gross income of more than US\$1.5 million are now required to put 0.5–2% of that income into science, technology or innovation. The amount of spending is determined by the type of company. Hydrocarbon companies are supposed to spend 2%; mining and electricity companies 1%; and all other industries 0.5%, the website SciDev.Net reported last week.

The money can stay in the company's science programmes, or be redirected to universities or other centres.

## Grand Canyon book survives flood of protest

Despite assurances, the US National Park Service has yet to review whether or not it should sell a book that gives a religious explanation for the formation of the Grand Canyon.

In 2003, bookshops run by the park service started to sell the book, *Grand Canyon: A Different View*, which asserts that the Grand Canyon was formed about 6,000 years ago by a flood during Earth's creation. Geologists protested, saying that scientific data contradict the creationist explanation. In late 2003, the park service said that it would review the matter (see *Nature* 427, 186; 2004).

But according to park service spokesman David Barna, the book remains on sale. And *The New York Times* reported on 5 January that the park service never completed its promised review. "It is not our role to tell people what to believe," Barna explained in a written statement.



Anthrax mailed to Senator Tom Daschle's office caused widespread immune responses.



# BUSINESS

## Reinventing an industry

Two years after a radical change that brought India's patent laws into line with international trading rules, the country's drug makers are taking a new direction. **Apoorva Mandavilli** reports.

**K**iran Mazumdar-Shaw is India's uncrowned queen of biotechnology. She started her company, Biocon, in a garage in 1978 with just Rs10,000 (US\$225) in working capital and has built it into the country's largest biotech company, with 1,800 employees and revenues last year of \$180 million.

Said to be the richest woman in India, Mazumdar-Shaw was in the spotlight last September when her Bangalore-based company launched the first new drug to be developed, tested and taken through approval by an Indian company. The drug, BIOMAb-EGFR, is a monoclonal antibody for treating head and neck cancers.

This could be the harbinger of a brighter and more innovative future for India's drug industry, which until recently relied on supplying cheap 'generic' copies of drugs — many of which were still under patent elsewhere.

That all changed in January 2005, when India brought itself into compliance with the Trade-Related Aspects of Intellectual Property Rights (TRIPS) — international rules that forbid the copying of patented drugs.

The transition has gone smoothly. "Companies are playing by the rules," says Frederick Abbott, a professor of international law at Florida State University who knows the Indian drug industry well.

Few of India's estimated 20,000 drug companies have been driven out of business so far. But there's little sign that many are ready to follow the trail blazed by Biocon and develop patented products of their own. There are fewer than 100 new compounds in various stages of development in the entire Indian industry. Most companies lack the resources to do their own research and instead are fighting for tiny pieces of the generics market.

"Today every company has learned how to make generics and your business window is very small," says Mazumdar-Shaw. "I think it's very clear to all the pharma players that unless they have an innovation strategy, the way ahead is not going to be easy."

Some of Biocon's competitors like to point out that its much-trumpeted new drug isn't a genuine Indian original, based as it is on a

molecule licensed from the Cuban Centre for Molecular Immunology.

Other observers are more charitable, however. "It's a very big deal because they've developed the drug, and development is not something India is known for," says Raghu Cidambi, head of corporate intellectual property for Dr Reddy's Laboratories in Hyderabad, one of the nation's largest pharmaceutical companies. "I think it's a very positive sign for the industry as a whole."

### Looking outwards

The larger Indian companies such as Dr Reddy's and Delhi-based Ranbaxy, India's largest pharmaceutical company, are working hard to reinvent themselves in light of the TRIPS agreement. They are spending up to 10% of revenues on research and development, acquiring companies abroad and forging collaborations with foreign partners.

This small group of industry leaders sees several factors working in their favour. Basic drug research in India costs much less than it does in the West, and the full cost of developing and approving a drug in India could be just \$100 million — compared with up to \$1 billion in the United States. India has the largest number of US Food and Drug Administration-approved drug manufacturing facilities of any nation outside the United States. Its universities churn out thousands of graduates — particularly chemists — each year, who

work for a fraction of Western salaries.

"There is a change now for the better," says Pradip Bhatnagar, Ranbaxy's vice-president for drug discovery. "There's nothing we don't have — knowledge, good people, infrastructure, experience." But he admits that genuine innovation remains rare. "Although the scientists are there, the inquisitiveness that's required for medicinal chemistry or drug discovery is not," Bhatnagar says.

Ranbaxy is one of the world's top 10 generics manufacturers, with 2006 sales of \$1.2 billion. Last year, the company spent about \$85 million on research and development, including a collaboration with Geneva-based Medicines for Malaria Venture to develop a synthetic form



Star turn: actor Shahrukh Khan (left) helps Biocon's Kiran Mazumdar-Shaw launch BIOMAb-EGFR.

of the antimalarial drug artemisinin.

"With our purchasing power here, it equals about \$400 million to \$500 million," says Bhatnagar. However, the investment in research has squeezed Ranbaxy's profit margins. Generics, Bhatnagar says, will always remain a part of the business.

### Trouble at home

Even there, Ranbaxy faces growing competition. Generic drugs are currently worth about \$40 billion worldwide and with so many best-selling drugs soon coming off patent, that is expected to double. Multinational generics producers such as Israeli company Teva Pharmaceutical Industries and Sandoz, a German subsidiary of the pharmaceutical giant Novartis, are expected to expand their Indian operations. "Indian companies have an advantage but it's marginal because everyone else is coming," says Cidambi.

Research-based pharmaceutical companies that had left India because of its lack of patent protection are also returning, often in alliance with local partners. American companies Bristol-Myers Squibb and Pfizer, as well as the British firm AstraZeneca, have signed contract-research deals with Biocon. UK-based GlaxoSmithKline has entered into a broader alliance with Ranbaxy, and Novartis from

**"There's nothing we don't have: knowledge, good people, infrastructure, experience."**  
— Pradip Bhatnagar



D. SARKAR/AFP/GETTY IMAGES

Switzerland and Denmark-based Novo Nordisk are collaborating with Dr Reddy's on developing drugs for diabetes.

In the short term, Indian firms need these partnerships to help run large clinical trials, position their drugs in the United States and Europe — and bring in money. Biocon has subsidiaries that do contract research in early drug development and help conduct clinical trials. These companies account for 15% of its revenues and 25% of its profits. "If we had built Biocon as a stand-alone company, it would have been very expensive and almost unaffordable," says Mazumdar-Shaw.

There are still some disincentives to foreign companies setting up in India. The Drug Controller General of India has been slow to approve new drugs and it still has arcane rules mandating, for example, that every drug be tested for its effect on male fertility. And although the law has been changed to protect patents, it doesn't shield trade secrets or proprietary information.

But as the competition increases, local companies will need to learn fast. "It's going to be a real battle," says S. V. Kapre, executive director of the Pune-based Serum Institute of India, the world's largest manufacturer of DPT (diphtheria, pertussis and tetanus) and measles vaccines. "We've always had the philosophy of making something from scratch," he adds. "We don't want to be bottlers and fillers for somebody else."

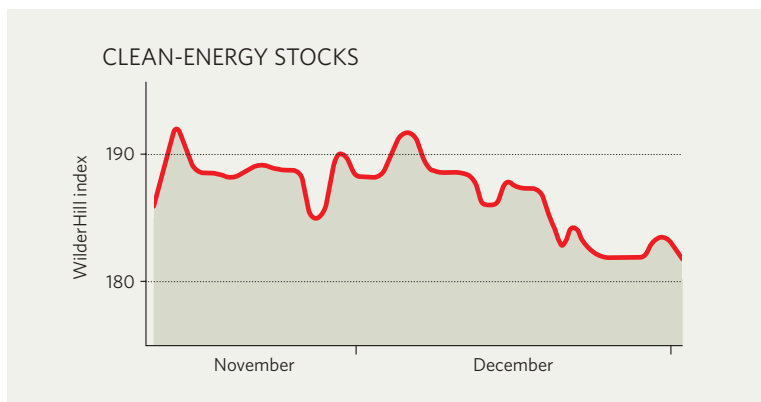
## IN BRIEF

**FACTORY PLAN** Plastics-based electronics hardware took a big step forward with the announcement that electronics company Plastic Logic, based in Cambridge, UK, has raised \$100 million to build a production plant that will make control circuitry for flexible, plastic displays. The company, founded in 2000 by University of Cambridge physicists, says it will build the plant in Dresden, Germany, by the end of 2008. The technology uses thin-film transistors made of semiconducting plastic substrates.

**VIAGRA VERDICT** A Beijing court has ruled in favour of Pfizer in a long-running case over counterfeit versions of its anti-impotence drug Viagra. The Beijing No. 1 Intermediate People's Court ruled on 28 December that two Chinese manufacturers, Beijing Health New Concept Pharmacy and Lianhuan Pharmaceuticals, should stop producing their version of the drug and that the latter should pay the New York-based drug company 300,000 yuan (\$38,400) in compensation. The case has been running since 2004, when China's patent review board ruled against Pfizer's patents, and the verdict is viewed as a bellwether of China's readiness to enforce foreign companies' drug patents (see *Nature* 440, 990; 2006).

**CASH AT THE READY** Codon Devices, a Cambridge, Massachusetts-based company that specializes in DNA engineering, successfully raised \$20 million in a second round of financing from venture capital. The new round was led by Highland Capital Partners, one of the United States' largest venture-capital firms. Codon was founded by researchers at the Massachusetts Institute of Technology in 2004, and is one of the first commercial companies to specialize in the emerging field of synthetic biology.

## MARKET WATCH



2006 drew to an unusually quiet close in the normally volatile clean-energy sector. With no big movements in the oil price the market ran fairly flat — as it had done for much of the second half of the year.

But that stability won't last long into 2007, predicts Rob Wilder, chief executive of San Diego consultancy WilderShares. His firm runs the WilderHill Clean Energy Index (see graph), which tracks the performance of a basket of US-listed firms with strong interests in energy sources not based on fossil fuels or nuclear power.

Wilder notes that the new, Democrat-controlled US Congress is likely to turn national attention towards both energy security and global warming in the new year. The forthcoming publication of the fourth report of the Intergovernmental Panel on Climate Change will doubtless

have the same effect — and could also help clean-energy stocks. There are risks, however: last week, the oil price began to tumble and that could scare some investors away from alternative-energy stocks. Either way, Wilder says, "it is highly unlikely that the sideways movement of the index will continue".

Michael Liebreich, chief executive of London-based New Energy Finance, agrees that there could be turbulence ahead for clean-energy stocks. The sector needs three things to happen for sustained growth during 2007, he says: investors not to overinflate prices; governments not to oversubsidize new energy sources; and the industry to build up its supply capacity. "These are the areas where we don't want to see any car crashes," Liebreich says.

Colin Macilwain



**B**onny Island's strategic location, at the mouth of the Niger River in west Africa, has always been a blessing and a curse. An important trading post for at least seven centuries, the island shipped hundreds of thousands of slaves from inland tribes to the Americas. Its Ibani people have retained good relations with foreign businesses ever since, even as the island's main export changed from people to oil and gas.

The traditional rulers of Bonny Island have helped to maintain a relative stability here in this oil-rich but desperately poor region. That is one reason why Bonny Island hosts the largest-ever industrial investment in Africa: a US\$15-billion gas-liquefaction plant that compresses and exports millions of tonnes of natural gas — enough to feed half the gas needs of France.

The plant, run by a consortium of oil and gas companies known as Nigerian Liquefied Natural Gas (NLNG), expanded spectacularly after Nigeria's government decided that the gas that had previously been flared off should be compressed and exported. The plant now draws thousands of skilled workers from all over the world, as well as impoverished workers from the surrounding regions. Although accurate figures are hard to come by, the island's population has grown from perhaps 30,000 a decade ago to 100,000 or more today.

That influx has set the stage for a new menace. A combination of migration, fast money and rampant prostitution may be placing the island on the brink of an AIDS explosion. And so NLNG, with help from its sister oil companies, has chosen Bonny Island for an experiment that could change the face of public health in the entire Niger Delta: a partnership between the local community, government agencies and oil and gas companies that

strives to study, test for and eventually control AIDS on the island.

The project is unusual in that it seeks to pull together industry and government money to support an initiative that will be built from the ground up, within the community of Bonny Island. It places great emphasis on self-help groups and other mechanisms that will involve all of the relevant groups in the

community, from schoolteachers to sex workers, in the nitty-gritty of AIDS education and prevention.

"We are in Bonny Island because it is isolated and relatively secure," says Donald de Korte, a veteran AIDS physician and consultant to the US drug giant Merck who is overseeing the project. Relatively is a key word here; in late December, the oil company Shell evacuated family members of expatriate staff from Bonny Island after car bombings and armed attacks elsewhere in the delta.

### Money for nothing

The project, called Ibani-se after the local people and their word for 'energy', is the first of its kind in Nigeria. The country has an HIV prevalence of about 4% — much lower than that in southern or eastern Africa, perhaps because of western Africa's retained tradition of male circumcision. Still, an estimated 3 million adults in Nigeria are HIV-positive — more than anywhere else except South Africa and India.

With the exception of a few big city hospitals, little counselling and treatment are available; only about 5% of Nigerians who need antiretroviral treatment are getting it, says de Korte. Nigeria is awash with oil and cash, but not much of that makes it into basic health services for the country. Oil money has helped

## ON THE BRINK

Containing a potential HIV explosion in the strife-torn Niger delta is a tough job — but circumstances are forcing the oil and gas industries to confront it. **Colin Macilwain** reports.



Testing times: Bonny Island's run-down hospital will be the medical focus of the Ibani-se initiative.





**Slick solutions: the owners of the gas-liquefaction plant on Bonny Island are driving plans to improve public health on the island.**

to fight AIDS, Tuberculosis and Malaria, or the US President's Emergency Plan For AIDS Relief (PEPFAR).

The key to the project is not to pour in drugs, money or equipment, which would in all likelihood be stolen, but rather to build something that is sustainable. To that end, de Korte and his team have painstakingly consulted a wide range of people on the island, including businessmen, sex workers, fishermen, the Navy, drivers, market traders, churches and mosques, and the twelve 'houses' of the Ibani people.

### Reaching high

The initial goals of Ibani-se are, at first glance, quite modest. The project plans to have 50 patients on antiretroviral treatment by the end of the year, for example, and to have 800 patients — a third of those who could benefit — on treatment by the end of 2009. But given current conditions — with minimal public-health infrastructure, few people being tested or counselled, and practically no one getting drugs — these goals are much more ambitious than they initially appear.

The medical focal point of the initiative will be Bonny Island's dilapidated, 60-bed general hospital, where a team of 26 medical staff deal with 450 patients a week. The hospital is supposed to be being renovated; last month, the patients were crowded into just one of the four wards while they waited for the other wards to be fixed. "A lot needs be done" to upgrade

the hospital, admits Douglas Pepple, the senior physician at the hospital, "but you have to start somewhere."

At present, the only people who can get AIDS drugs on Bonny Island are pregnant women, who are entitled to treatment until their baby is born. Everyone else has to go to a hospital in Port Harcourt — the sprawling capital of the delta, 50 kilometres up the pirate-ridden river. No one knows anyone on the island who has managed to complete the trip.

Ibani-se may help to change that situation. It was launched on 1 December 2006 — the last World AIDS Day. The day has a high, if ambiguous, profile in Nigeria. In the morning, local radio stations discuss the issue of

AIDS, and O Network, MTV's popular African affiliate, broadcasts a question-and-answer session in which South African schoolchildren interrogate ex-US president Bill Clinton about AIDS. It also shows videos of ditties such as *Put You On the Game*, in which Los Angeles rapper The Game threatens women with a life of prostitution.

Then, King Edward XI of Bonny Island and Peter Odili, the governor of Rivers State, arrive in the town of Akiama to officially launch Ibani-se. After a slow start the ceremony soon gathers steam, as about 500 schoolchildren and other islanders crowd into the community hall.

The king — who lives for most of the year in Nigeria's capital Abuja, where he chairs a confederation of Nigeria's traditional rulers — tells the crowd about an initial baseline study for the project. The survey, based on a representative sample of Bonny Island's

**"We need to rise up against the stigma and discrimination that surrounds AIDS."**

— King Edward XI

the Nigerian government to clear its \$32 billion debt since 2004. But the economic activity generated by the production of oil and gas has inflated prices in the Niger Delta, fuelling social division.

The Ibani-se initiative is being supported initially by industry sponsors including NLNG, Shell and ExxonMobil, who have done other outreach programmes on the island. There is a discernable dichotomy here. On the one hand, the initiative hopes to achieve a sustainable partnership that is anchored in the community. But others have a more jaundiced view — grounded in past experience in the region — that nothing will function on Bonny Island unless NLNG does it, and pays for it.

Still, the project has a determined leader in de Korte, a Dutch physician who used to run Merck's business in Africa and then, from 2000 to 2004, led one of the continent's largest and most comprehensive AIDS treatment programmes, in Botswana. De Korte was tapped by officials from Shell and NLNG to chair the panel overseeing the initiative, after they heard him talk about his views on how AIDS should be fought. Since then de Korte has been working to constitute Ibani-se as a non-governmental organization (NGO), which would make it eligible for financial support from international bodies such as the Global Fund



Spreading the word: despite signs advising people of the risks, many are not aware of how real they are.



adult population, found that an estimated 9,600 people on the island are now HIV-positive. "There's an urgent need to commence an HIV treatment programme," the king says. "We need to rise up against the stigma and discrimination that surrounds AIDS."

**"The community is yet to come to terms with the fact that AIDS is a big issue here."**  
— Samson Sunday

### Empty gestures

For the event, Shell had shipped in a \$40,000 machine to test CD4 counts — a measure of immune function — and dropped it, incongruously, on the floor of the hall. The gesture does not exactly thrill Ibani-se's organisers; nobody on the island knows how to use the machine. Thanking Shell for the gift, the king takes an adroit swipe at the donor: "\$20,000 [sic] is nothing compared with what we will do have to do in the future to mount a project that must be sustainable," he tells the community.

The budget of the programme has been modest so far, with about \$600,000 donated by NLNG last year and \$1.4 million set to be forthcoming in 2007. Merck has contributed the time of de Korte and other outside consultants, and plans to eventually supply drugs to the programme at a reduced rate.

The initiative's small group of permanent staff are Nigerian healthcare workers, hired mainly from other NGOs. "The community is yet to come to terms with the fact that AIDS is a big issue here," says one of the team, Samson Sunday. "They know it exists, but they see it as a blight from the outside that only affects a small number of people. The first promising sign is that community leaders are accepting that they are sitting on a time bomb if they don't do something soon."

The project's start-up has been much tougher than anticipated. Ibani-se still hasn't attained the legal status that it seeks as an NGO, despite trying for the past year and a half. Basic logistics are immensely problematic. Project managers didn't want new vehicles at first, as ostentatious transportation is the first thing that tends to divide community projects in Africa from the communities they purport to serve. But as a result, the project has become too dependent on NLNG for logistical support. Ibani-se's staff have found, to their frustration, that while they wait for transport and phone lines, staff at the gas company have other priorities.

The headquarters of the project is also still located on NLNG's fortified residential compound, whose manicured lawns and smart buildings look more like a retirement complex in Florida than the 'real' island outside the gates. Ibani-se is planning to move to one of

the only concrete buildings in Bonny Town, the island's largest village. The building was constructed 20 years ago by Shell as a community library, but it has never been used and has stood derelict ever since.

It is palpable that local participants in the project are expecting some material gain. "There is hardly any volunteerism on the island," says de

Korte, who wants as many people as possible on Bonny Island to devote some of their own time to the project. But so far, "community engagement has only scratched the surface". The whole exercise has been much more difficult than expected, he admits. "I underestimated the Nigerian environment."

### Platforms for success

On the other hand, de Korte is proud of the baseline survey done last summer by a Nigerian NGO, the Society for Family Health. The survey forms the basis for the project's action plan and provides the first detailed examination of sexual practices and HIV prevalence on Bonny Island. It shows that the prevalence of HIV is now 7.8% — double what it was five years ago. The study is not yet complete or cleared for release, but it is already identifying areas that Ibani-se could focus on, such as the sexual behaviour of the local sex workers and the 17,000 staff and contractors at the NLNG plant.

Not everyone on the island was surveyed, however. An extension to the liquefied gas plant, being built by an international consortium led by the US corporation Halliburton, employs another 1,700 workers. The consortium declined to participate in the survey, citing security concerns. Unlike the gas liquefaction plant itself — which is increasingly staffed by Nigerians — almost all the workers in the extension are white expatriates.



Brian Buckley, general manager for production at NLNG, says that the Ibani-se project is just one component of the gas company's plan to tackle social issues on the island. He says that he is hopeful that Halliburton workers will participate in the project later on. NLNG staff have already done so, he points out, although not many of the expatriates volunteered for testing. From the company's point of view, Buckley says, the Ibani-se programme is vital to the long-term future of the plant. "We don't want to end up with our workforce depleted, like the mines in South Africa," he says. "The kids growing up on this island are the future of this community."

World AIDS Day may have provided a glimpse of that future. Before and after the king's ceremony, long lines of people queued patiently to be counselled individually in a small interview cubicle, then give blood and be tested for AIDS. The four Ibani-se counselors worked for hours and tested 100 people, but eventually ran out of time. They took down details of 121 more people for subsequent testing — taking the project some way towards its initial goal of testing 3,000 people by the end of the year.

If Ibani-se takes off as planned, its cost will rise significantly: thousands of islanders could end up on antiretroviral drugs priced at perhaps \$60 per month. De Korte is confident that





**Fighting fit:** the Ibani-se project aims to protect the future of the children on Bonny Island.

the Global Fund and other sources will help to foot the bill. But a spokesman for the fund, Jon Liden, says that public-private partnerships are “not as easy in practice as we had thought”. Efforts to fight AIDS in Nigeria, he says, have varied tremendously in quality; on several occasions, the fund has not renewed grants to Nigeria, “due to poor performance”.

### Addressing the stigma

World AIDS Day saw the project off to an auspicious start though, and the king's words provided some assurance that the initiative has buy-in at the right levels. “We are pleased that the king has given his support,” says Ibiba Chidi, a project manager for Ibani-se.

For now, one of the project's focal points will be education. Stigma continues to surround AIDS on Bonny Island: health workers have uncovered very few people who admit to being HIV-positive. At the ceremony, a woman who was billed as speaking for people who are HIV-positive said she was, in fact, speaking out for a friend who is.

The other focal point, again reflecting data from the baseline survey, will be the activities of sex workers on the island. Project officials estimate that some 500 prostitutes work in Monkey village, on the main drag between the

NLNG residential compound and the gas plant, where men gather at night to meet girls.

25-year old Joy Odudu runs a makeshift wooden bar in the ramshackle village. “I came to the island to work, but was looking for a job for a year and none was coming,” she says. She claims that site managers wanted a bribe of 40,000 naira (more than US\$300) from women seeking employment there.

Odudu, whose arm was shot by river bandits, says that about half the men who frequent her bar after dark are expatriates, and that some of them will pay girls extra not to use condoms. She says she doesn't know how often this happens — “we can't follow them into their rooms” — but that arguments between clients and girls, and between girls, have exposed the practice. One Australian worker, she asserts, was paying girls \$300 or \$400 for nights of unprotected sex.

A nightly curfew, introduced after three oil workers were kidnapped on Bonny Island last August, is hurting business, Odudu says. But the expatriates are still coming, curfew or no curfew, says Peters Kunamon, headman of a

section of Monkey village, Odudu's protector and a member of Ibani-se's community interface committee.

Asked about the project itself, both Odudu and Kunamon say it will win their support if it provides them with some material benefit. “It is supposed to be voluntary — but we need to eat,” says Kunamon, who also runs a pharmacy store in the village that will sell condoms. “The project is good if it gives us something.” Kunamon mentions something else that others are not keen to point out: women with day jobs on the site, he alleges, are also ‘working’ there at night.

Lucy, a girl at the bar, is just as direct. As she realizes that visitors are here for a chat, not for sex, her mood swings between charm and irritation. “You want jiggy-jiggy? Condoms or no condoms — what does it matter?” she asks sarcastically, voice rising. “We have to pay 100 naira here for water, to wash in the morning. We need to eat! We have nothing!”

### Setting the scene

Back inside the compound at the close of World AIDS Day, the Sun goes down soon after six, as it does all year round so near the equator. Soon after, as the curfew takes effect, engineers converge on the open-air cocktail bar, where local musicians provide some outdoor entertainment. It's an attractive setting, with elegant log cabins and decks and an outdoor pool whose racing section is 49 metres long. If it were an olympian 50 metres, apparently, the law would require NLNG to open it to the public; in this intensely fortified zone, that wouldn't do at all.

Buckley is there, on what happens to be his last day as production manager. Next week, he is off to run an even larger (and more peaceable) liquefied-gas facility in Oman. He relaxes with a beer and a group of friends, most of whom sport polo shirts and chinos, the international uniform of the professional man at rest.

But all around the outer reaches of the extensive bar pavilion, smartly dressed Nigerian girls who don't look at all like petrochemical engineers are hanging out. Their teeth and eyes sparkle in the evening light, they have the great posture

you only get from learning to balance ten litres of water on your head at the age of four, and they're keen to strike up conversation with strangers. “Five thousand naira,” one of them whispers, by way of introduction. “All night long!”

Colin Macilwain writes for *Nature* from Edinburgh.

**“We don't want to end up with our workforce depleted, like the mines in South Africa.”**  
— Brian Buckley



Life is full of events that are basically games, from paying for a meal to bidding in an auction. Can incorporating a quantum strategy into the rule book increase your chances of winning? **Navroz Patel** reports.

# States of play

**T**he year is 2015. Representatives from the world's leading defence firms are gathered in a high-security room at the Pentagon. Each is seated facing a screened-off console. Together, they are bidding in an auction that will allocate the various contracts for a \$100-billion project to develop a new breed of fighter aircraft. But this is no ordinary auction — it is designed to ensure that the companies bidding will collectively offer the US government the lowest price for the whole project.

In an ordinary auction, one firm might be prepared to offer a low price for a particular contract if it could be sure that another company, with which it has collaborated, will win the contract for another component. If this situation could be encouraged, then the government could shave billions off the cost of the project. But such conditional bids are unlikely as none of the firms wants to reveal its bidding strategy to the others, or even to the auctioneer.

Which is where the new strategy comes in — the contracts for this project will be decided by a 'quantum auction'. The participants use their consoles to manipulate qubits (the bits inside their quantum computers) to produce quantum states that correspond to each of their desired bids. Because the bids are encoded in fragile quantum states, no one, not even the auctioneer, can read the information in the qubits without destroying them.

The success of the auction depends on repeated exchange of information between the bidders, through a quantum algorithm that operates on the qubits, until an optimal outcome is reached. In the process of reading this outcome, the individual bids are destroyed. This guarantees that losing bids will never be revealed and so dramatically increases trust between the participants. The ability to express conditional bids in a secure way is what leads to the more ideal outcome.

Think this all sounds too futuristic to take seriously? Think again. A group of researchers at HP Labs, the research arm of Hewlett-Packard in Palo Alto, California, is working to create the quantum protocols that could allow such an auction to take place within a decade. And their work is helping to bridge the gap between theory and practice in both game theory and quantum information processing.

The field of quantum game theory (QGT) is still in its infancy: just tens of economists and quantum physicists have so far published a total of 200 or so papers on the topic. Its roots stretch back to a paper published in 1999 by mathematician David Meyer of the University of California, San Diego<sup>1</sup>. In it, Meyer showed how a quantum approach always beats a classical strategy in a simple game where two players flip a coin. This is because the laws of quantum mechanics allow the coin to exist in a state that is a combination of heads-up and tails-up at the same time, so the person playing by classical rules will always be outmanoeuvred.

That's all fine in theory, but subsequent work on QGT has struggled to be taken seriously. "Much of the claimed superiority of quantum games to their classical counterparts has been the result of incorrect comparison," Meyer acknowledges. Some of the scepticism arises because both game theory and quantum information processing have unresolved issues. Like QGT itself, the field of quantum information processing is mostly theoretical and, because computers that can handle hundreds of qubits are several decades away, critics argue that much of the research is simply impractical. Game theory — which is used to analyse strategic situations and the behaviour of participants seeking to maximize their success in such games — also has problems, not least of which is the fact that some of its predictions contradict what happens in the real world. But by linking these two problem-

atic fields into QGT, researchers hope to find practical applications that can resolve issues in both fields.

For example, the prediction that rational players in many kinds of games will pursue their own selfish interests — referred to as defecting — is not borne out in real-life: people are often altruistic. In the game known as the prisoners' dilemma, if two players both cooperate, they earn more than if they both defect; but an individual who defects always gets a higher pay-off, regardless of what the other player does. Game theory predicts that rational players will all defect<sup>2</sup>, but in practice they often cooperate.

## Jail-break

At HP Labs, a team including experimental economist Kay-Yut Chen and quantum physicists Tad Hogg and Raymond Beausoleil is making theoretical and experimental advances in QGT that may address some of these criticisms. One advantage of studying QGT over other applications of quantum technology is that fewer qubits should be needed to play quantum games, so it might be possible to implement them fairly soon. HP Labs hopes that Chen and Hogg's work on quantum auctions will ultimately lead to a new business model for selling digital content on the Internet that might help to discourage illegal downloading.

In terms of game theory, how digital content is sold and distributed on the Internet can be seen as a 'public-goods game'. Much as in the prisoners' dilemma, which is a two-player version of this type of game, selfless choices by members of a group reap greater benefits for the group as a whole, although selfish behaviour produces the greatest personal gain. So how do content providers prevent illegal downloading of software, music and videos, when others have already paid for the privilege of exclusive access?

The problem is similar to that faced by a table of diners in a steakhouse when it comes to settling the bill. Experience shows that when people agree to split the cost of a meal between



**"We can safely assume that 99% of the population could not accurately be described as quantum physicists."**  
— Kay-Yut Chen



IMAGESTATE/ALAMY

quantum physicists,” says Chen. “So we needed to understand how players would behave in this kind of game.”

In 2004, Chen and Hogg set out to show that it is possible for people to play primitive quantum public-goods games<sup>4</sup>, using a dozen or so students from Stanford University. The experiments, which mimicked both classical and quantum versions of the games, did not involve actual quantum computers or qubits, instead the students interacted with computers that simulated the players’ operations on quantum states. Indeed, the students were completely unaware that any feature of the game aped the laws of quantum mechanics.

### Place your bets

In both the classical and quantum games, the students were each given \$100 and randomly grouped into pairs to play a version of the prisoners’ dilemma in which the choice was either to keep the \$100 or to contribute it to a common fund. The pair were better off if they both cooperated and contributed to the fund, but individuals got higher pay-offs by defecting and not contributing no matter what the other player decided.

In the quantum version of the game, the decision of whether or not to contribute was determined by a more complex process intended to simulate quantum entanglement between qubits assigned to each player. Entanglement means that the quantum states of two or more disparate objects are intrinsically linked in some way — so changing the state of one object automatically affects the state of the others. Players were asked to pick three numbers that controlled the probabilities of the various outcomes of the entanglement; before and during the game, they were given software tools to help them to understand the consequences of their choices and those of their opponents. But because their opponents’ choices remained unknown, and because the probability of an outcome was rarely 100%, each player could only guess at the best numbers to choose.

Overall, the students cooperated roughly 50% of the time in the quantum games, as opposed to just 33% of the time in the classical version. “We were surprised. Our lab experiments

them, if the group is small — just a few people — the diners tend to order more modestly priced items from the menu, and may even offer to pay more than their share of the bill. But as the size of the group increases, people feel a greater sense of anonymity and may conclude that their individual decisions will have a smaller impact on the overall bill. The end result is defection: some people will order the most expensive steak rather than a cheap cheeseburger and, to add insult to injury, might not even stump up their share of the bill. Game theorists refer to this kind of player behaviour, which is detrimental to the wider good, as free-riding.

### Game on

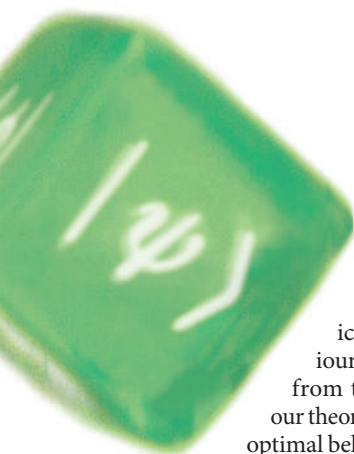
In their quantum approach to the public-goods problem<sup>3</sup>, Hogg and his colleagues focused on multiple games of prisoners’ dilemma played between pairs of people within a larger group. In a three-player group, for example, each player would play two games, one with each of the other two. Unlike classical game theory, the quantum version of the prisoners’ dilemma predicts that players will cooperate in 50% of the games played. And, unlike in real-life clas-

sical situations, Hogg and Chen predicted that the greater the number of players involved, the less of a problem free-riding should be.

In a classical prisoners’ dilemma, defecting is always the best strategy for rational players. In the quantum version of the game, which intrinsically links the two players’ decisions through quantum mechanics, the outcome is often probabilistic in that the players can’t be sure their pay-off will be the same each time, even when all players make the same choices. This uncertainty alters the structure of the game to one in which no single choice is best. “The quantum version of the game is more like scissors, paper, stone,” says Hogg. In such games, there is no single move that will win each time, so a mixed strategy — choosing moves randomly or based on guesswork — is the best option for rational players.

This analysis worried the researchers as people don’t usually play mixed-strategy games very well. As with all game-theory research, one question loomed large: will the outcomes predicted by Hogg and Chen occur in reality? “We can safely assume that 99% of the population could not accurately be described as





showed that players without any training in quantum mechanics exhibited behaviour indistinguishable from that predicted by our theory," Hogg says. This optimal behaviour seemed to emerge because players tried to second-guess what their opponent's strategy was, resulting in a mixed strategy even though that was not what they intended.

### Joint effort

Although these experiments involved two-player games, and so simulated two-way entanglement between players' qubits, Hogg and Chen also studied larger groups of three or four players, in which each permutation of pairs played a game of prisoners' dilemma. The results suggested that free-riding decreases as game size increases to three or four players. If confirmed with larger groups, this effect would be highly desirable in the context of Internet piracy, where the number of players, that is downloaders, can run into tens of millions. "There are many complexities in this context, but it falls within the scope of what quantum public-goods games research can address," Hogg says.

Hogg and Chen believe that their experiments can mimic more complex entanglement within larger groups of players, using relatively few qubits. "The way we are creating these protocols — with the small number of qubits required and pair-wise entanglement — is, we believe, physically doable within five to ten years," says Chen. Such experiments might allow them to study economic problems in which the arguments for players using a quantum approach are compelling. Chen says they hope that their work on quantum auctions will soon produce a protocol that will enable collaborative auctions without needing to rely on trusted third parties (the auctioneer) or prior agreements to reach a preferred outcome.

Knowing what your competitors will bid is a major uncertainty at every auction, but knowing what a potential collaborator will bid and factoring that into your strategy is a more subtle uncertainty, referred to by economists as



How we split a restaurant bill mirrors a game whose outcome could be improved by quantum mechanics.

'allocative externality'. In the fighter-jet example, this might translate to a company that is bidding to build the undercarriage wanting to bid less if another specific firm wins the contract for the landing gear. "Allocative externality is a prevalent problem in the awarding of large government contracts, and is something that all non-quantum auctions fail to address," says Chen.

Last summer, Chen and Hogg began to investigate how people behave in a quantum auction — although they first chose to study a simpler highest-bid-wins auction without collaborative bids. In this, the bidders decided how much they wanted to bid, and then deployed a quantum protocol similar to that used in the public-goods game to encode their bid and keep it private.

**"Much of the claimed superiority of quantum games to their classical counterparts has been the result of incorrect comparison."**

— David Meyer

To process the multiple bids, the researchers simulated the action of a quantum algorithm designed to find the maximum value of a quantum state — in this case, the maximum revenue received by the auctioneer. In an actual implementation, this search would be performed during the repeated exchange of qubits between the auctioneer and participants until an optimal answer was converged upon. At that point the auctioneer would perform a measurement of the qubits to decide the outcome of the auction, and all the losing bids would be destroyed.

Initial three-player auctions revealed flaws in the quantum search algorithm. "Deficiencies in the algorithm enabled players to potentially

misrepresent bids so that ultimately they could win with a lower bid," says Hogg. He adds that the team is currently redesigning the algorithm to avoid such problems before tackling more complex collaborative auctions.

Chen and Hogg hope to introduce allocative externality into their quantum auctions in the next year or so. Although it will be some years before they begin to experiment with actual qubits, Meyer is encouraged by the team's experiments: "I'm fairly optimistic about the kind of research they are doing." Given the obvious security advantages of quantum auctions, Meyer believes that there is now a greater chance that QGT will prove to be genuinely useful.

Hogg and Chen's studies have also been welcomed by physicist Paul Ellsmore, co-author of a recent UK report on the commercial prospects for quantum information processing<sup>5</sup> and chief executive of semiconductor firm Nanion in Oxford. Conceptually, quantum games have some attractive features, Ellsmore says, but the creation of robust algorithms to encode quantum information has been lacking. "Commercialization had been considered a dirty word in this field," he says. "It's encouraging that researchers are now developing algorithms that could be implemented with technology that will be viable in the near future."

**Navroz Patel is a freelance writer based in New York City.**

1. Meyer, D. A. *Phys. Rev. Lett.* **82**, 1052–1055 (1999).
2. Nash, J. F. *Proc. Natl Acad. Sci. USA* **36**, 48–49 (1950).
3. Chen, K.-Y., Hogg, T. & Beausoleil, R. *Quant. Inf. Process.* **1**, 449–469 (2003).
4. Chen, K.-Y. & Hogg, T. *Quant. Inf. Process.* **5**, 43–67 (2006).
5. Corker, D., Ellsmore, P., Abdullah, F. & Howlett, I. *Commercial Prospects for Quantum Information Processing* (Quantum Information Processing Interdisciplinary Research Collaboration, 2005); available at [www.qipirc.org/files/Commercial%20Prospects%20for%20QIP%20v1.pdf](http://www.qipirc.org/files/Commercial%20Prospects%20for%20QIP%20v1.pdf)

## Hidden subsidies and risks pump up real price of oil

SIR — Congratulations on your excellent Biofuels Special (*Nature* **444**, 654 and 669–678; 2006). Not only is this an important topic for scientists and environmentalists, but also developing energy products that move the United States away from its over-dependence on oil is a critical national-security priority.

There are many exciting possibilities for expanding renewable fuels in the United States and around the world. Today's ethanol from corn and sugar will be complemented by cellulosic ethanol from many energy crops, as well as by next-generation biodiesel and perhaps other renewable fuels that have yet to be developed. Moreover, these will complement liquid fuels made from abundant coal reserves.

Although I agree that markets need to determine which fuels emerge as the most competitive, we should recognize that these new fuels are not competing on a level playing field with oil. The true costs of oil to our economy and national security are much higher than the price paid at the pump. Much of the world's oil is controlled by authoritarian regimes that increasingly use their oil resources for political gain. In addition, the price we pay for oil does not cover the risks to the economy from delivery disruptions, price spikes, environmental hazards and the vast cost of protecting supply routes and infrastructure. Milton R. Copulos, president of the National Defense Council Foundation, testified to the Senate Foreign Relations Committee last year that the minimum military subsidy Americans pay to protect Middle East oil trade routes is \$50 billion a year (see [www.senate.gov/foreign/hearings/2006/hr060330a.html](http://www.senate.gov/foreign/hearings/2006/hr060330a.html)).

Consequently, Congress is debating a variable price floor for alternative fuels so that oil producers cannot use temporary price cuts to undercut the new fuels that make America more secure. Such a floor should allow all fuels to compete, so the best ones may emerge.

**Dick Lugar**

United States Senate, 306 Hart Senate Office Building, Washington DC 20510, USA

## Poaching laws are useless without solid enforcement

SIR — Your News Snapshot of tigers in India's Nagarhole National Park ("A night out in the park" *Nature* **444**, 413; 2006) quotes a conservationist's view that tiger conservation is more at risk from loss of habitat than from the tiger trade.

This may be true in a few protected areas

such as Nagarhole, where anti-poaching laws protecting tigers and their prey are strictly enforced. However, only 23% of the priority landscapes where tigers survive today are protected areas (E. Dinerstein *et al.* *Setting Priorities for the Conservation and Recovery of Wild Tigers: 2005–2015: A User's Guide* WWF, WCS, Smithsonian and NFWF-STF, 2006), and enforcement in many of those places is poor. In 2005, for example, poachers were found to have wiped out the entire tiger population of Sariska Tiger Reserve in India (D. Banks and B. Wright *Skimming the Cat: Crime and Politics of the Big Cat Skin Trade* Environmental Investigation Agency/Wildlife Protection Society of India, 2006).

If we want to save wild tigers, we will have to reduce demand for tiger parts and crack down on illegal trafficking throughout Asia. We also have to improve on-the-ground efforts beyond reserve boundaries to achieve real landscape-level conservation outcomes.

**Brian Gratwicke**

Save The Tiger Fund,  
1120 Connecticut Avenue NW, Suite 900,  
Washington DC 20036, USA

## Turks fighting back against anti-evolution forces

SIR — Your recent Special Report "Anti-evolutionists raise their profile in Europe" (*Nature* **444**, 406–407; 2006) draws attention to the strong anti-evolution climate in Turkey, and concludes pessimistically. However, the teaching of evolution is not a lost cause in Turkey.

It is true that the situation is grave. In a recent survey of acceptance of evolution, Turkey scored the worst among 25 countries, with less than 30% of the population accepting it (J. D. Miller *et al.* *Science* **313**, 765–766; 2006).

The major reason for this has been the conservative influence on education in Turkey during the past 25 years. In 1985, the then minister of education contacted creationists in the United States, a cooperation that led to the inclusion of creationism in the high-school biology curriculum and textbooks.

Furthermore, anti-evolution views are not restricted to textbooks. In a study conducted by one of us in the capital, Ankara, last year, only 47% of the 147 biology and science teachers surveyed said they accepted evolution. More disturbing is that it was accepted by significantly fewer of the young teachers and by only 26% of the 257 14-year-old students.

On the other hand, Turkish scientists have been striving to reverse this trend. It has been publicly criticized by the Turkish Academy of Sciences ([www.natcensci.org/resources/](http://www.natcensci.org/resources/)

[rncse\\_content/vol19/8371\\_cloning\\_creationism\\_in\\_turkey\\_12\\_30\\_1899.asp](http://rncse_content/vol19/8371_cloning_creationism_in_turkey_12_30_1899.asp)). A group of graduate students known as Evrim Caliskanlari, or 'hard-workers for evolution', has started translating the University of California, Berkeley's Understanding Evolution website into Turkish (see [www.universitekonseyleri.org/evrimcaliskanlari](http://www.universitekonseyleri.org/evrimcaliskanlari)).

Most forcefully, a non-governmental association, Universite Konseyleri Dernegi, has filed a lawsuit against the Ministry of Education, demanding that creationism should be removed from textbooks and evolutionary biology should be covered appropriately in the curriculum.

The ministry has responded by asserting that darwinism is scientifically suspect — using publications by the US intelligent-design Discovery Institute for reference. It goes on to claim that developed countries are including creation-like theories in their curricula and to imply that evolution is not compatible with Turkish 'culture and values'. At this point it is unclear how the case will turn out.

If more Turkish scientists showed their support for the efforts that are being made, and put pressure on their academic bodies to take a pro-evolutionary position, this would certainly influence both the ministry and public opinion. Better late than never.

**Mehmet Somel\***, **Rahsan Nazli Ozturkler Somel†**, **Aykut Kence‡**

\*Max Planck Institute for Evolutionary Anthropology, Deutscher Platz 6,  
D-04103 Leipzig, Germany

†Department of Education Science,  
Helmut Schmidt University, Holstenhofweg 85,  
22043 Hamburg, Germany

‡Department of Biology,  
Middle East Technical University,  
06531 Ankara, Turkey

## Readers see red over low-impact graphics

SIR — *Nature* authors are increasingly using colour within figures. May I remind them and the graphics editors that a significant proportion of their readers, including me, cannot distinguish red from green? Issue number 7120, for example, was not atypical in having six figures that used these two colours. The impact of their figures — and just possibly the impact factor of the journal — will be improved if all the audience can see all the data.

**Chris Miall**

School of Psychology, University of Birmingham,  
Birmingham B15 2TT, UK

**Contributions to Correspondence may be submitted to [corres@nature.com](mailto:corres@nature.com). They should be no longer than 500 words, and ideally shorter.**



## BOOKS &amp; ARTS

# The supreme problem solver

An exploration of the life and science of physicist Hans Bethe.

## Hans Bethe and His Physics

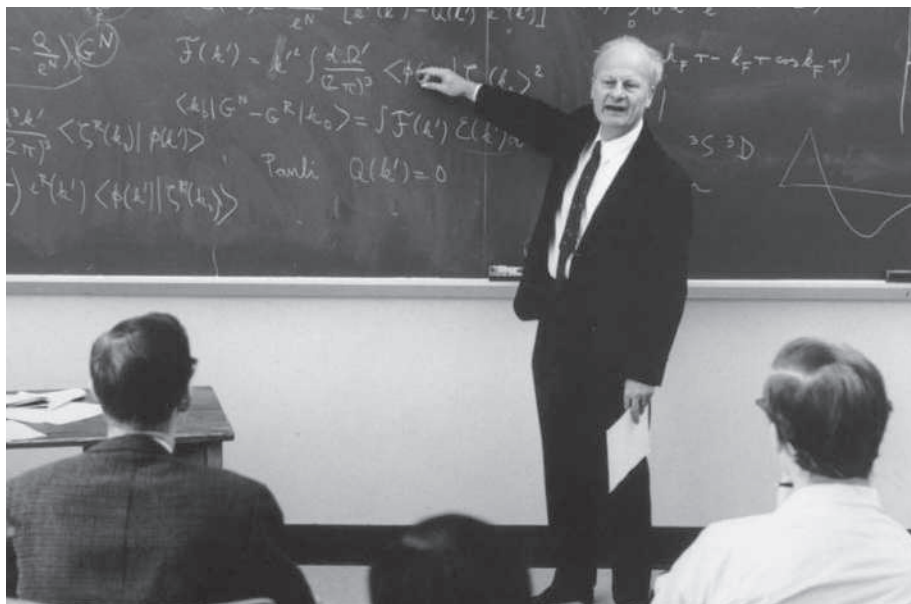
edited by Gerald E. Brown & Chang-Hwan Lee

World Scientific Publishing: 2006. 314 pp.  
\$98, £56 (hbk); \$38, £22 (pbk)

## David Wark

It is almost 20 years since the winter's day when I walked out of West Bridge at the California Institute of Technology from my thesis defence, with blood only slightly diluting the adrenaline in my veins, to deliver my newly approved opus to the registrar and record my PhD. As I swaggered across the concrete arch spanning the fountain in the centre of the campus, an elderly, heavy-set man was slowly making his way up the other side, his face a study in concentration as he peered into the water. I wanted to trigger a conversation to somehow mark my newly fledged status, but the words caught in my throat and I stepped aside. Even in my euphoric state I knew better than to interrupt Hans Bethe when he was thinking. Six years of my effort had produced a pedestrian but passable thesis, whereas an afternoon of Bethe's thinking could transform an entire field of physics.

This awe of Bethe and his work is universal among those who knew him, and the contributors to *Hans Bethe and His Physics* are certainly no exception. Freeman Dyson called him "the supreme problem solver of the twentieth century", and John Bahcall said it looked as if Bethe's output was the result of a conspiracy by several people all signing their work with the same name. His output must be nearly unmatched in duration, too — he published his first paper (with his father) at the age of 18 in 1924, and his last was submitted by his co-authors to the preprint servers six months after his death in March 2005. Along the way he picked up the 1967 Nobel Prize in Physics for work in the 1930s on the energy-production mechanisms of stars; he introduced the Bethe ansatz, which has found applications throughout physics and mathematics; he wrote three long articles for *Reviews of Modern Physics* in 1936–7 that were dubbed the 'Bethe Bible' and are said to contain everything that anybody else knew about nuclear physics and quite a bit more besides; and he gave his name to some now-essential formulae for calculating nuclear masses, the energy loss of charged particles passing through matter, and much more as well.



CORNELL UNIV.

Hans Bethe won the physics Nobel prize for work on stars, but he has influenced a wide range of fields.

But it was his 1947 calculation of the Lamb shift that perhaps best shows the power of his direct approach to physics. At a conference on Long Island, Willis Lamb announced that accurate measurements of the fine structure of atomic hydrogen were in conflict with Paul Dirac's prevailing theory of electrodynamics. Many of those present, including such towering figures as Niels Bohr and Robert Oppenheimer, saw this as the demonstration of a profound theoretical crisis that would require revolutionary new developments to resolve. On leaving the conference, Bethe took up the problem and had managed to calculate a value of this 'Lamb shift' that matched the observations before his train had travelled the 75 miles to his destination. He did this not with any revolutionary developments, but by performing the simplest calculation that he thought might match the data. This was the Bethe way, or as he put it: "Learn advanced mathematics in case you need it, but use only the minimum necessary for any particular problem," and "be prepared to make conjectures if it helps". All real problems are solved with approximations, so genius in physics often consists of knowing which mistakes to make, and Bethe was a master of it. His calculation convinced others that the fledgling field of quantum electrodynamics (QED) was on the right track, and helped spur the developments that led to the great success

of QED and hence much of modern theoretical physics. His extremely pragmatic attitude to calculation, which he passed on to the legion of students he taught in his 71-year career at Cornell, still influences the field today (although looking at the current state of theoretical physics, perhaps not enough).

As head of the theory division at Los Alamos during the Manhattan Project, Bethe played a key role in constructing the first nuclear weapons. A refugee from Hitler's Germany, he was well aware of the danger that nuclear weapons could pose in the hands of totalitarian regimes. So, like many who had fled fascism, he devoted himself to ensuring that the democracies got these awful weapons first. After the war, Bethe was sensible enough to know that the development of nuclear weapons would continue in the United States whatever he did. He therefore continued his involvement in the belief that he could be a more effective voice for moderation from the inside — although quotes such as "I sometimes wish I was a more consistent idealist" show his ambivalence about such work. He was also an authoritative and effective advocate for arms control and international cooperation.

To fully describe such a formidable list of accomplishments will keep many a scholar busy for the foreseeable future. *Hans Bethe and His Physics* is much too short to even attempt

such a task. It grew out of a request by Bethe to Gerry Brown to explain his physics to the world; Brown and Chang-Hwan Lee edited this book. It makes no claim to be a comprehensive biography — it is more like a taster menu drawn from half-a-dozen different restaurants. Like Bethe's work, the chapters range from extremely detailed physics to non-technical consideration of some of the biggest issues facing humanity. Much fascinating insight into some of the key figures can be gained without any scientific background, but some chapters require a knowledge of physics to degree level or beyond. A few of the later articles have a faint echo of axes being ground, but a book with no trace of controversy would be a poor reflection on Bethe's life.

I would especially recommend the book to

anyone who has been involved in any way with the events described, as it brings alive many of the physicists that some of us knew only slightly, or by reputation alone. Reading about Bethe's pure pleasure as he bulldozed problem after problem using the simplest tools he could get away with was an inspiration — I wanted to grab a piece of paper and have a go myself. He kept that pleasure to the end. A colleague who gave a seminar at Cornell in the week of Bethe's death told me that Bethe closely followed the talk and obviously enjoyed it. His last words to a younger physicist, quoted in one of the best and most touching articles in the book, were simply "Carry on". We will, Hans.

David Wark is at Imperial College London and the Rutherford Appleton Laboratory, Chilton, Didcot, Oxfordshire OX11 0QX, UK.

fleshy ornaments can be condition dependent, and so provide potential mates with reliable information about quality.

It is surprising that until only a decade ago, the study of avian coloration proceeded in ignorance of colour science. A mini-revolution has since occurred in how colour is considered, measured and analysed in behavioural, ecological and evolutionary studies of birds. This change is reflected in *Bird Coloration*. Previously, almost all studies of avian coloration were made on the assumption that birds see colours in much the same way as humans. This assumption was flawed. We now know that many, if not most, diurnal birds are sensitive to ultraviolet wavelengths, to which humans are blind. Moreover, colour vision in birds arises from four classes of retinal cone cell, whereas humans have only three, probably giving many birds an extra dimension to their colour vision. In addition, experiments have shown that ultraviolet reflectance and illumination invisible to humans influences birds' decision-making in several contexts, including mate choice, foraging and provisioning chicks.

The two volumes of *Bird Coloration* provide an excellent up-to-date overview of the topic. Despite having 20 contributors (including many prominent people in the field), the book is generally well focused, although the editing could have been tighter in places.

The science of bird coloration today is genuinely interdisciplinary, drawing on behaviour, ecology, pigment chemistry, the physics of light, the microstructure of reflectors, physiology, immunology, molecular genetics and vision research. Most of these areas are well represented in the book. I particularly enjoyed the chapters on topics that have had relatively little attention in the past few years, such as the numerous non-sexually selected functions of avian coloration. These also provide a reminder of the many issues that remain to be unravelled. Consider, for example, even well-studied species, such as the zebra finch or the blue tit. There are still no convincing explanations for their combinations of

## A bird's-eye view

### **Bird Coloration: Vol. 1, Mechanisms and Measurement; Vol. 2, Function and Evolution**

edited by Geoffrey E. Hill & Kevin J. McGraw

Harvard University Press: 2006.

Vol. 1, 544 pp, £59.95; Vol. 2, 496 pp, £59.95

#### **Andrew T. D. Bennett**

Who can fail to be beguiled by the extravagant plumage colours of birds of paradise, peacocks and parrots? But there's more to feathers than showy displays. These remarkable, lightweight structures, which evolved 150 million years ago, perform a range of functions, both within individuals and across species. For example, they are one of the best thermal insulators known; they have several aerodynamic properties, creating refined aerofoil structures, producing lift and reducing drag; they can be waterproof; they are resistant to wear and abrasion, and are self-repairing; they perform a wide variety of sensory functions; and they create humid habitats for parasites and symbionts. Finally — and this is the focus of *Bird Coloration*, a two-volume work edited by Geoffrey Hill and Kevin McGraw — feathers produce a plethora of colours that demand an evolutionary explanation.

Simply put, an animal's coloration results from a trade-off between crypsis, for protection from predators, and gaudiness, for mating advantages. But the whole story is much more fascinating, and understanding it requires a firm grasp of both sexual-selection theory and colour vision. Sexual selection is a term coined by Darwin to explain the evolution of traits — often extravagant ones, such as the peacock's tail — that appear to confer disadvantages in natural selection but offer mating advantages. It was largely ignored for more than a century, but over the past three decades it has been one of the most intellectually vibrant areas

of evolutionary biology. Its powerful body of theory now provides evolutionary explanations for a vast range of behaviour, ecology and life-history variation. For example, in humans it can explain mating patterns and even sporting and musical aptitude.

The test-bed for much of the work on sexual selection has been birds and their coloration. Why? Birds have grabbed biologists' attention in a way that nematodes have not, so there is a vast knowledge base on which to draw. In addition, the gaudy coloration and extravagant displays of birds require explanation and are amenable to testing. Vision plays a key role too — indeed, birds have perhaps the most advanced visual system of any vertebrate.

An explosion of interest was fuelled in the early 1980s by Bill Hamilton and Marlene Zuk's remarkable theory that much bird coloration evolves from sexual selection on traits that indicate condition, particularly disease status and heritable resistance to parasites. Many studies have since showed that feathers and



**Putting on a show:** a peacock displays its colourful tail.

G. DOUWMA/GETTY IMAGES



colours and their proximity and areas.

Colour science is an important but complex subject. Those studying behaviour and ecology are now taking this on board, as several chapters demonstrate. How are receptor signal-to-noise ratios dealt with, and does the Vorobyev and Osorio model best describe most avian colour spaces? What are the advantages and disadvantages of the different approaches to analysing reflectance and illumination spectra, and how do these vary with the different level and type of biological questions posed? How

precisely are different cone inputs combined and weighted? What is the taxonomic distribution of spectral sensitivity in birds? How are photonic crystals genetically coded? These are all areas of current intrigue, and quite a few are well covered here.

A sound understanding of colour vision, then, underpins our understanding of bird coloration. This can be a knotty area, with most hard evidence coming from colour-mixing experiments, retinal physiology and sequence data, so contributors specializing in these areas

would have been welcome. However, bird coloration is a huge topic and anyone organizing an overview should be heartily congratulated.

Another overview can be found in a special issue of *American Naturalist* due out this month. Both publications offer much for those studying coloration in other taxa too, because birds provide model systems for investigating animal coloration. Besides, birds are just so exquisitely beautiful.

Andrew T. D. Bennett is at the School of Biological Sciences, University of Bristol, Bristol BS8 1UG, UK.

## A cultural state of mind

### The Culture of Our Discontent: Beyond the Medical Model of Mental Illness

by Meredith F. Small

Joseph Henry Press: 2006. 224 pp. \$27.95

#### Lewis Wolpert

In medieval times, madness was regarded as a moral perversion or even satanic possession. There has since been a continuous increase in the use of drugs, from alcohol to antidepressants. The current culture of using pharmaceuticals to treat mental illness concerns the anthropologist Meredith Small. In *The Culture of Our Discontent*, she examines how such illnesses are viewed in different cultures, and whether, for example, shamans and sorcerers might be able to suggest useful treatments. Much of the book is based on interviews with psychiatrists and anthropologists, and it provides a novel approach to mental illness.

However, Small underestimates how successful talking therapies, such as cognitive therapy, have become. In my view, psychoanalysis, by contrast, is not far removed from witchcraft. One of the problems facing talking therapies is that it is very difficult to do a double-blind clinical trial.

A relatively new approach to psychiatry is based on darwinian evolution. For example, depression is both disabling and common, and has a significant genetic component, so why has it not been selected out? Might depression have a useful function? Could it be a cry for help, or prevent people from pursuing unattainable goals? Most plausible is that sadness is an essential and adaptive human attribute that has become abnormal, much as cancer cells are normal cells that have become malignant. Similar questions relate to schizophrenia. Does it give someone the advantage of not needing social attention, for example? I don't find this persuasive, however. More relevant, I suggest, is the view that it involves key cognitive processes, such as insight and empathy, that have become abnormal.

Small describes how Harry Harlow used monkeys in the 1950s to try and understand depression. He found that infant monkeys separated from their mothers showed behaviour



Cultures such as the Aymara in South America have different views of mental illness to the West.

that could be thought of as depression. There was also a strong genetic component. However, they recovered when placed with other young and affectionate monkeys.

Central to the book is the extent to which culture influences the types and incidence of mental illness. Unfortunately, little good evidence is provided in relation to Western illnesses in other cultures, probably because they are not easy to diagnose. There is a widely accepted claim, for example, that the incidence of schizophrenia is 1% in many different cultures. However, depression is rare in the cold wastes of Iceland, and the number of teenage suicides is growing in Colombia.

Culture, like genes, is handed down and affects behaviour. It also determines what is considered normal. In one Native American community in California, seizures were respected, and in Siberia, many shamans are neurotic. In Botswana, an anthropologist saw a woman suddenly rip off her clothes and run into the bush. Quite slowly, a man went after her and they returned calmly. The locals explained that sometimes people had to do that. In southern China, koro, the male fear of losing the penis, can reach epidemic proportions. By contrast,

anorexia is entirely a Western condition and may be related to affluence. Agoraphobia is also Western — there is less need to run for cover in the jungle.

In China there were no acknowledged cases of depression under Mao, but Arthur Kleinman found that patients suffering from neurasthenia — aches and pains — had the characteristics of depression. Such somatization is also common in depressed patients in the West, but this is not explored here.

As Small points out, there are no reliable data on the diagnoses or remedies for different mental illnesses in different cultures, but in the non-Western world there is usually someone to blame, often guided by a witch. Small notes that in Western culture we do not tend to believe in curses, voodoo or magic, but what about religion and much of alternative medicine?

The examples of mental illness in different cultures raise many interesting problems and require further research. We also badly need to improve our understanding of the biology of mental illness, and then the cultural influences may become clearer.

Lewis Wolpert is at University College London, Gower Street, London WC1E 6BT, UK.

A. POHL/STILL PICTURES

# The great detective

## The Memoirs of Sherlock Holmes: The Adventure of Silver Blaze

by Arthur Conan Doyle

G. Newnes: 1894

### Richard Gregory

Nostalgia precedes memory. We experience this paradox in the Sherlock Holmes stories by Arthur Conan Doyle. They certainly transport me to the London of hansom cabs and dark cobbled streets, a world that vanished a generation before my eyes opened — indeed before I had eyes, or an I, to accept and solve the problems of seeing. The eye is the detector, the brain the great detective, reading clues from the eyes and the other senses.

The importance of small but significant clues for seeing was suggested to me years ago by Sherlock Holmes. It soon extended to science, my father being an astronomer, and it then steered my experiments on visual perception for half a century. My father measured the distances to stars. I tried to measure and understand distortions of visual space. Eyes cannot signal the presence of objects to the brain, but only shapes and colours and movements, as clues to what objects are like and what they can do. Surely perceptions are hypotheses of what might be out there — the closest we can get to reality.

Clues tend to be small and inconspicuous, but may actually be nothing at all. For Silver Blaze, a racehorse stolen from his stable at night in *The Memoirs of Sherlock Holmes*, it was nothing that gave the game away:

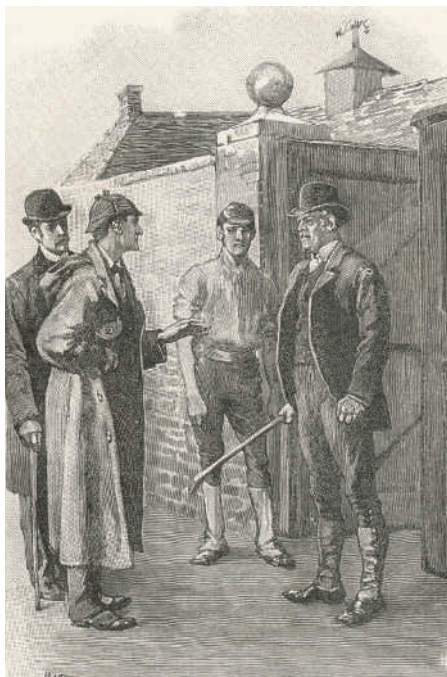
*"Is there any other point to which you would wish to draw my attention?"*  
*"To the curious incident of the dog in the night-time."*  
*"The dog did nothing in the night-time."*  
*"That was the curious incident," remarked Sherlock Holmes.*

A nothing can be a clue for seeing, as when a nearby object is inferred from the gap it makes in the background. Surprising gaps can evoke illusions of familiar objects. This is very different from the tradition of needing physical stimuli to evoke responses and perceptions. A surprising non-event, or non-object, may revise the brain's hypothesis, changing our perception.

Appreciating a clue depends on its context, either in time and place or within a theory. In the pockets of the Silver Blaze suspect were found a collection of

seemingly ordinary objects. Which of these were clues for finding out who was the Silver Blaze villain? Some of them were hard to accept as significant without the context of a theory.

At first, only an ivory-handled knife with a delicate, inflexible blade stood out as significant, as it was an unusual object. It was an important clue, but so too were some ordinary matches and a candle. On Holmes' working theory, they pointed to the villain nobbling the horse, with a delicate operation that was practised in the dark on nearby sheep. Holmes confessed that this was a long shot:



*"It struck me that so astute a man as Straker would not undertake this delicate tendon-nicking without a little practice. What could he practice on? My eyes fell on the sheep..."*

When the clues were explained to the owner of Silver Blaze, he could see:

*"I have been blind!" cried the Colonel. "Of course, that was why he needed the candle and struck the match."*

For events, or non-events, to be clues, knowledge and imagination are needed. If someone is said to be 'clueless', this does not mean there are no clues, but that the knowledge and imagination to appreciate them are lacking. We see this in the contrasting character of the police officer, Inspector Gregory (no relation!).

*"Inspector Gregory, to whom the case has been committed, is an extremely competent officer. Were he but gifted with imagination he might rise to great heights in his profession," said Holmes. "It is the one quality which Gregory lacks."*

The inspector had checked the surrounding ground for tracks of the horse but found none. Holmes and Watson checked again, and Holmes soon found what they sought:

*The track of a horse was plainly outlined in the soft earth in front of him, and the shoe which he took from his pocket exactly fitted the impression. "See the value of imagination," said Holmes. "It is the one quality which Gregory lacks. We imagined what might have happened, and acted upon the supposition, and find ourselves to be justified. Let us proceed."*

Holmes' imagination was effective, being based on knowledge. His assistant Dr Watson makes explicit Holmes' inferences, or makes them conscious:

*"There was no wind that night, I understand," said Holmes.*  
*"None; but very heavy rain."*  
*"In that case the overcoat was not blown against the furze bush, but placed there."*

Holmes' inferences jump beyond the evidence to dangerous hypotheses. Misleading murder clues have hanged innocent men, and misleading visual clues produce illusions. Yet jumping beyond the evidence is essential for discovery and invention. Sherlock Holmes and science are both justified, even though sometimes inevitably wrong.

We started with a paradox: nostalgia can precede memory. For writers such as Conan Doyle who can take us far away and to a time long gone, the solution is simple — we see with their eyes. But after a pipe or two, Holmes might have said: "Not quite so simple, Watson. Eyes can detect only the features of the here and now, as clues for the 'great detective' to infer the world of objects as hypotheses that we accept as reality." Watson would have replied: "You mean, the brain is a Sherlock Holmes machine."

Then, smiling gently, Watson asked: "What is consciousness?" Lifting his violin to play unearthly music, Holmes might have replied: "I haven't a clue."

**Richard L. Gregory is in the Department of Experimental Psychology, University of Bristol, Bristol BS8 1TU, UK.**



## ADAPTIVE IMMUNITY

# Care for the community

A memory-based immune system may have evolved in vertebrates because of the need to recognize and manage complex communities of beneficial microbes.

**Margaret McFall-Ngai**

All vertebrates have a type of immunity known as adaptive immunity, which allows them to respond to each fresh encounter with the microbial world on the basis of past interactions. Invertebrates, however, rely entirely on the innate immune system, an ancient mechanism present in all animals that does not typically 'remember' previous encounters. The remarkable memory component of the vertebrate adaptive system has been thought to provide heightened resistance to microbial pathogens. But this assumption presents a conundrum — invertebrates are no less challenged by the microbial world, nor are they less able to remain healthy.

How have invertebrates been so successful — they comprise more than 96% of animal species — without the benefit of an adaptive immune system? Many have thought the answer lies with lifestyle: because invertebrates are small, bear many young and are short-lived, they should have no need for a memory-based immune system that is suited to the long haul. But numerous invertebrates are large, have only a single offspring each year and live a long life. The quahog clam, for example, remains healthy for up to 250 years, despite pumping bacteria-rich sediments across its internal organs. Clearly, 'long-life' lifestyle features can evolve in the absence of an adaptive immune system. I propose a different explanation: that adaptive immunity has evolved in part to recognize and manage complex communities of beneficial microbes living on or in vertebrates.

Recent advances in molecular biology that allow biologists to quantify and identify an animal's normal microbiota suggest that the presence of complex communities of coevolved bacteria is a shared feature among vertebrates. In general, the coevolved partnerships of invertebrates seem to be much less diverse.

For example, characterization of human microbial partners has identified more than 2,000 bacterial species that typically associate as communities inside us — including on or in our skin, guts and mouth. These coevolved, resident communities are often in direct contact with our tissues, are relatively resistant to perturbations, such as starvation, and provide us with the metabolic benefit of millions of additional genes and activities. By contrast,

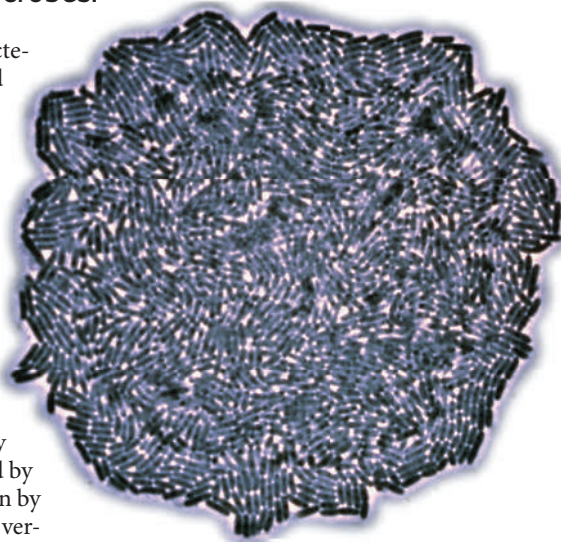
fewer than 100 species of human bacterial pathogen have been identified, and exposure to these is generally rare and transient.

Meanwhile, evidence is emerging that most of the associations between invertebrates and the microbial world are different. Although an invertebrate gut tract can harbour dozens of bacterial species at any time, recent studies have identified only a small number (fewer than eight) of resident bacterial species. Thus, the gut microbiota of invertebrates is numerically dominated by tourists, with a composition dictated by the external environment rather than by stable residents, as is characteristic of vertebrates. Under these conditions, invertebrates might use their innate immune system to limit the range of interactions with the microbial world, treating all but a few species as unwelcome guests. By contrast, the evolution of the vertebrate immune system is likely to be strongly affected by the need to maintain a substantial resident microbiota.

The mechanisms by which invertebrates recognize and manage those resident microbial partners that they do have are unknown, but there are at least three possible strategies. First, there are examples in all major phyla of invertebrates that have beneficial bacterial associations with between one and three species that are maintained intracellularly, and so are invisible to the innate immune system. Although widespread among invertebrates, such associations seem to be absent from vertebrates.

A second strategy is to maintain a physical barrier between the host tissue and the microbes. The unusually complex, highly stable consortia harboured in the hindguts of termites and their relatives are effectively separated from the host tissues by a layer of chitin that precludes direct host–microbe interaction.

Third, even without a dedicated memory function, the innate immune system may be capable of managing the simple communities of resident bacteria typically present in the invertebrate gut. Recently, the number of specific recognition components of the innate system has been found to be much higher in some invertebrates than earlier estimates suggested; these specific recognition elements might provide a mechanism for keeping track of a



**Group effort: did microbial communities inspire the need for adaptive immunity?**

number of microbial partners. The absence of these three strategies among vertebrates is consistent with the idea that the adaptive immune system has provided them with a different, more versatile microbial-management strategy.

The advances in technology that have inspired these ideas should also allow us to test their predictions. Careful characterization of the gut microbiota of various vertebrates and invertebrates could address the basic premise that all vertebrates have a coevolved microbiota, whereas invertebrates rarely do. Similarly, comparative physiology could test the prediction that maintenance by vertebrates of coevolved microbial consortia, both throughout life and across generations, provides advantages, such as more efficient digestion, that are not available to invertebrates. Finally, there is the question of whether there is a price to be paid for the permissiveness of the adaptive immune system. Is autoimmunity a collateral consequence of the bargain we've cut with microbes? ■

**Margaret McFall-Ngai is in the Symbiosis Cluster and the Department of Medical Microbiology and Immunology, University of Wisconsin, Madison, Wisconsin 53706, USA.**

#### FURTHER READING

Hedrick, S. M. *Immunity* **21**, 607–615 (2004).  
Weng, L. et al. *Genome Res.* **16**, 316–322 (2006).  
Dethlefsen, L., Eckburg, P. B., Bik, E. M. & Relman, D. A. *Trends Ecol. Evol.* **21**, 517–523 (2006).  
Backhed, F., Ley, R. E., Sonnenburg, J. L., Peterson, D. A. & Gordon, J. I. *Science* **307**, 1915–1920 (2005).

EYE OF SCIENCE/SPL

CONCEPTS

## PALAEOONTOLOGY

# Embryonic identity crisis

Philip C. J. Donoghue

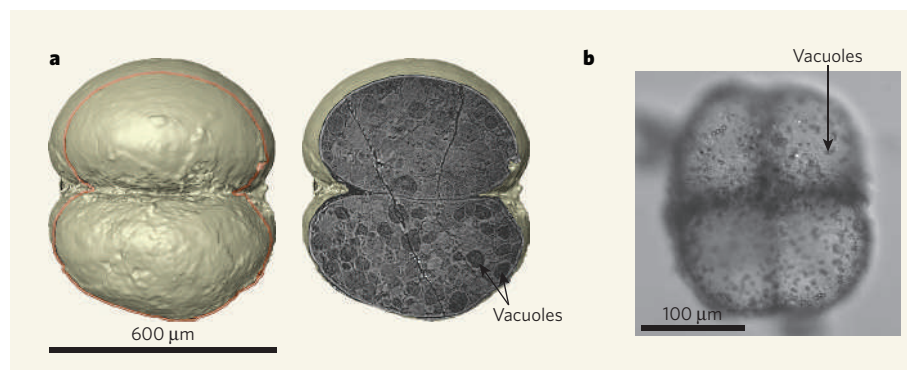
**The oldest known animal fossils, identified as eggs and embryos, had been expected to reveal secrets from a period of great evolutionary change. Will the latest theory about the fossils' origins confound these hopes?**

The origin of animals is almost as much a mystery as the origin of life itself. An abundant fossil record extends back 542 million years to the beginning of the Cambrian period, testifying to the establishment of all of today's main groups of animals by this time. However, the degree to which animal evolutionary history extends beyond the Cambrian is a controversy rich in speculation but sparse in evidence. It is no wonder that the 1998 report of fossilized animal embryos from the Doushantuo phosphorite rocks of southern China created a stir<sup>1</sup>. At more than 580 million years old, these were the oldest unequivocal animal fossils, and, as embryos, they provided a glimpse of animal embryology at a time when today's main animal groups were emerging. These fossils have revealed ancient patterns of cell division and cell arrangement<sup>2</sup>, and promised to reveal further secrets about developmental evolution at this crucial time. But this prospect may have been dashed by a study on page 198 of this issue<sup>3</sup> presenting a compelling reinterpretation of these fossils — not as animal embryos, but as giant bacteria\*.

The fossils in question are species of *Parapandorina*, which were thought to be embryos, and *Megasphaera*, which were proposed to be fertilized eggs. Most *Parapandorina* fossils were preserved only in their very earliest stages of embryonic development, and were composed of 2, 4, 8, 16, 32 or 64 cells, with some examples having undergone further rounds of cell division — a process known in embryos as cleavage. These remarkable fossils are preserved in calcium phosphate, and are present in such abundance that they are often the main constituent of the rock. But the icing on the cake is that, in the Doushantuo rocks, quantity goes hand in hand with quality. The detail of preservation can be staggering, with features such as cell nuclei and subcellular vacuoles — membrane-bound compartments — being observed<sup>2</sup>.

The embryos were originally identified as colonies of green algae, but were later classified as animal embryos because of their comparatively large size (typically just under a millimetre in diameter), and because the cell

\*This article and the paper concerned<sup>3</sup> were published online on 20 December 2006.



**Figure 1 | Fossil embryos?** **a**, *Parapandorina* fossils from the Doushantuo region of southern China have features akin to those of animal embryos. This image of a two-cell *Parapandorina* 'embryo' was derived using X-ray microtomography. The virtual cross-section reveals the component cells separated by walls. The cells' interiors are dominated by what seem to be membrane-bound vacuoles or yolk particles. (Images courtesy of P. C. J. Donoghue.) **b**, Bailey *et al.*<sup>3</sup> suggest that the fossils could be bacteria similar to *Thiomargarita* (pictured), which exhibit some of the same features as animal embryos. The bacterium shown has divided into four cells, in which the vacuoles are clearly visible. (Image from ref. 5.)

walls show evidence of distortion in response to the division of their neighbours, suggesting that they are not rigid, as algal cell walls would be<sup>1,4</sup>. More specific evolutionary relationships to sponges and arthropods have been mooted, but definitive evidence has been lacking.

Bailey *et al.*<sup>3</sup> now show that, although *Parapandorina* and *Megasphaera* may not be algae, the features thought to be indicative of animal embryos could just as readily be those of bacteria. This is based on comparisons with *Thiomargarita* — giant sulphur bacteria that live in seafloor sediments along the Namibian coast and in the Gulf of Mexico. For instance, the pattern of cell division in *Parapandorina* is conventionally interpreted as cleavage. But this pattern is also seen in *Thiomargarita*, where it represents successive rounds of cell division. This results in clusters of two, four and eight cells<sup>5</sup> with geometries identical to those of *Parapandorina*<sup>3</sup>. *Thiomargarita* is also as large as the Doushantuo fossils, and, like *Parapandorina*<sup>2</sup>, its cells are densely vacuolated<sup>5,6</sup> (Fig. 1).

Intriguingly, *Thiomargarita* is capable of controlling phosphate mineral precipitation<sup>6</sup>, so Bailey and colleagues' theory<sup>3</sup> provides an obvious source for the phosphate that is so abundant in the Doushantuo Formation and that preserves the fossils themselves. Together

with evidence that *Thiomargarita* rarely progresses beyond a few rounds of cell division, this goes a long way to explain why fossils with few cells are so abundant in the Doushantuo phosphorites, and why adult specimens grown from the 'embryos' are apparently absent. Such issues are difficult to explain with the animal-embryo interpretation.

But equally nagging problems remain with the bacterial explanation of *Parapandorina* and *Megasphaera*. *Thiomargarita* has not been shown to undergo more than three rounds of cell division, whereas specimens of *Parapandorina* can be composed of more than a hundred cells<sup>2</sup>, suggesting at least seven cleavage events. Furthermore, cells of *Parapandorina* and *Megasphaera* cluster together in an enveloping membrane<sup>4</sup>; although cells of some species of *Thiomargarita* also cluster in a sheath, the species that best simulates the early cleavage stages of animals does not<sup>5</sup>. And although densely packed vacuoles are a feature of the Doushantuo fossils and the sulphur bacteria, in *Parapandorina* they are present throughout the cells<sup>2</sup>, whereas in *Thiomargarita* a dense layer of vacuoles is restricted to a thin cytoplasm surrounding one large, central vacuole<sup>5,6</sup>. Finally, bacteria don't possess cell nuclei, so the observation of preserved nuclei in *Parapandorina* is



hardly compatible with Bailey and colleagues' proposal.

No matter. Such quibbles do not diminish the central message of the authors' report, which is that, like all other theories about Precambrian animals, the classification of these fossils is far from resolved, even at the kingdom level. More data and critical analysis of the Doushantuo biota are required, such as that already provided by X-ray microtomography<sup>2,7</sup>. Only then can we assess whether any of its fossils address such overarching questions as the timing and embryological basis of animal origins. ■

Philip C. J. Donoghue is in the Department of Earth Sciences, University of Bristol, Wills Memorial Building, Queen's Road, Bristol BS8 1RJ, UK.

e-mail: phil.donoghue@bristol.ac.uk

1. Xiao, S., Zhang, Y. & Knoll, A. H. *Nature* **391**, 553–558 (1998).
2. Hagadorn, J. W. *et al. Science* **314**, 291–294 (2006).
3. Bailey, J. V., Joye, S. B., Kalanetra, K. M., Flood, B. E. & Corsetti, F. A. *Nature* **445**, 198–201 (2007).
4. Xiao, S. & Knoll, A. H. *J. Paleontol.* **74**, 767–788 (2000).
5. Kalanetra, K. M., Joye, S. B., Sunseri, N. R. & Nelson, D. C. *Environ. Microbiol.* **7**, 1451–1460 (2005).
6. Schulz, H. N. & Schulz, H. D. *Science* **307**, 416–418 (2005).
7. Donoghue, P. C. J. *et al. Nature* **442**, 680–683 (2006).

## PARTICLE PHYSICS

# Hard-core revelations

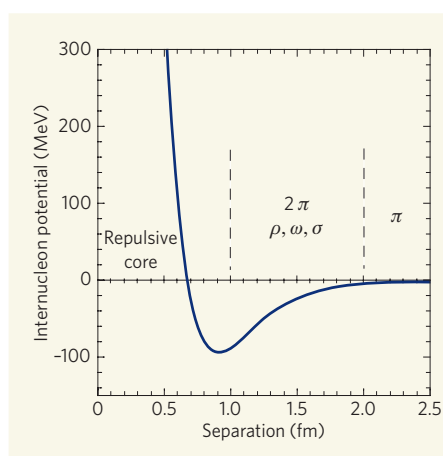
Frank Wilczek

**Our description of how the atomic nucleus holds together has up to now been entirely empirical. Arduous calculations starting from the theory of the strong nuclear force provide a new way into matter's hard core.**

Our quest to understand the force that holds atomic nuclei together has turned out to be a glorious adventure. Along the way we have found quarks, the coloured gluons that mediate the strong nuclear force, and a wonderful theory — quantum chromodynamics, or QCD. This theory has guided experimental research at the high-energy frontier, inspired dreams of 'unified field theories' that would embrace all nature's forces, and allowed theoretical physics to penetrate into the cosmology of the early Universe. In all this, the original problem of understanding nuclear forces has rather fallen by the wayside. That changes with what may come to be seen as a landmark paper by Ishii, Aoki and Hatsuda that has recently appeared on the arXiv preprint server<sup>1</sup>.

Ironically, from the perspective of QCD, the foundations of nuclear physics appear distinctly unsound. Famously, nuclear physics is best understood by modelling atomic nuclei as assemblages of protons and neutrons moving at much less than the speed of light. Yet QCD tells us that protons and neutrons are themselves built from quarks and gluons that move at very nearly the speed of light. These more basic particles carry colour charges, leading to the additional requirement that they be confined within 'bags' whose contents are overall colour-neutral. So far, so good: we understand, at least roughly, how an unbalanced colour charge produces a growing cloud of virtual particles (the process known as vacuum polarization), which has to be neutralized. The neutral cluster holds together as a unit, like marbles in a bag.

But why don't the separate proton and neutron bags in a complex nucleus merge into one common bag? On the face of it, the one-bag arrangement has a lot going for it. It would allow quarks and gluons free access to a larger



**Figure 1 | The nucleon-nucleon potential.** At distances of a few fermi, the force between two nucleons is weakly attractive, indicated by a negative potential. According to Hideki Yukawa's model<sup>2</sup>, this force is mediated by the exchange of particles known as mesons. The  $\pi$ -meson, or pion, the lightest of the mesons, accounts for the attractive force at the largest distances where it is felt, whereas heavier mesons ( $\rho$ ,  $\omega$ ,  $\sigma$ ) take over closer in. The picture changes abruptly, however, below a separation of just under 1 fermi. Here the force becomes strongly repulsive, preventing nucleons merging. Ishii *et al.*<sup>1</sup> provide the first theoretical calculations from quantum chromodynamics, the theory of the strong force, that reproduce the empirical form of this potential.

region of space, and so save on the energetic cost of localizing their quantum-mechanical wavefunctions. But in such a merger, protons and neutrons would lose their individual identities, and our traditional, quite successful model of atomic nuclei would crumble. What prevents that calamity?

We gain insight into this question by

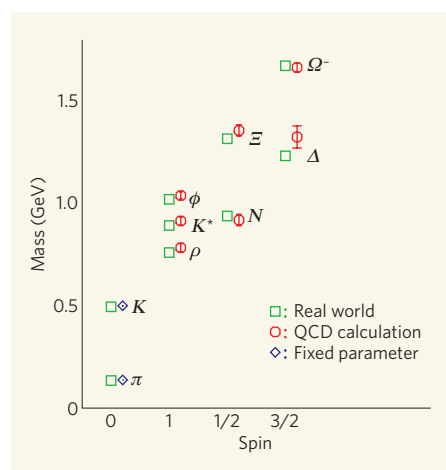
approaching it from the bottom up. Assuming that nucleons (protons and neutrons) are the appropriate starting point, what properties lead them to bind into atomic nuclei but to shun more intimate mergers? At an empirical level, the answer has been known for decades. The strong (that is, non-electromagnetic) internucleon force becomes significant at distances below a few fermi (1 fermi is  $10^{-13}$  centimetres). It remains attractive down to about one fermi, but, at shorter distances, very strong repulsion abruptly sets in (Fig. 1). In atomic nuclei, nucleons arrange themselves close enough to take advantage of the attraction, but they stay away from the notorious hard core. In particular, they do not merge.

This empirical 'answer' serves only to frame more questions. Does the fundamental theory produce a force like that? If so, why? As originally proposed by Hideki Yukawa<sup>2</sup>, the longest-range part of the strong internucleon force can be attributed to exchange of the lightest strongly interacting particles, known as  $\pi$ -mesons or pions. At shorter distances, exchanges of heavier mesons become important. As we approach hard-core distances, however, this meson-exchange picture becomes both unwieldy and dubious as the number of relevant mesons grows and their internal structure becomes resolved. Thus the existence of the hard core, which is absolutely crucial to the structure of matter as we know it, appears as a brute fact, opaque to theoretical analysis.

In principle, the equations of QCD contain all the physics of strong internucleon forces. But in practice, it is extremely difficult to solve the equations and calculate those forces. Ishii and colleagues' breakthrough calculation<sup>1</sup> required sophisticated algorithms, running on the biggest and fastest massively parallel computers currently available.

Why are the calculations so difficult? The main reason is simply that nucleons are complicated objects. It is often said that protons (and neutrons) are made from three quarks. That statement contains a kernel of truth, but it is a gross oversimplification. The kernel of truth is that a proton has the same conserved quantum numbers — charge and spin — as three quarks: two up ( $u$ ) quarks and one down ( $d$ ) quark,  $uud$ . Because these conserved quantities match, you can produce a proton by introducing three quarks and letting them settle down into a stable state of low energy.

In the process of settling down, however, the bare quarks dress themselves with a host of additional gluons and quark-antiquark pairs. Thus the wavefunction for a proton contains components with different numbers of gluons and quark-antiquark pairs, in addition to the basic three 'valence' quarks. To do that wavefunction justice, many different components must be sampled, and within each component the spatial distribution of its constituent quarks, antiquarks and gluons must be computed. The quantum-mechanical proton can contain all those configurations simultaneously. Existing



**Figure 2 | Massive success.** The fundamental equations of quantum chromodynamics contain three fixed parameters: the sum of the  $u$  and  $d$  quark masses, giving the pion mass; the mass of the  $s$  quark, which together with the  $u + d$  mass sets the kaon mass ( $K$ ); and an overall coupling strength fixed using the masses of mesons that contain heavier quarks (not shown). Given these, the masses of many observed particles have been computed with the precision indicated by the error bars. Notably here,  $N$  stands for nucleon (proton and neutron). Thus we have a first-principles account of the origin of the most of the mass of ordinary matter.

computers, based on classical physics, cannot. (For those familiar with the jargon: classical computers do not support superposition and entanglement.) The configurations must be laboriously constructed one by one, stored, and correlated. It is a job for teraflop machines running at full steam for months.

In recent years, first-principles calculations of the masses of strongly interacting particles — collectively known as hadrons and including the nucleons — have reached precisions of 5% or less, that precision being limited by computer power. This fundamental account of the origin of most of the mass of ordinary matter surely ranks as one of the greatest scientific achievements (Fig. 2). But yesterday's sensation is tomorrow's calibration, and the frontier of numerical QCD is set to expand into more difficult terrain. To study the nuclear force one must, of course, create two nucleons from scratch, and then study how the total energy depends on the distance between them.

That is exactly what Ishii *et al.*<sup>1</sup> have done. Their numerical results convincingly demonstrate that hard-core repulsion is a consequence of QCD: the strong internucleon force exhibits weak long-range attraction and strong short-range repulsion. Limitations of computer power did require the authors to use  $u$  and  $d$  quark masses slightly larger than the real-world values, so the magnitude and range of the hard core that they calculate should be taken as indicative, but not definitive.

To do justice to existing data, and to validate the method, much more work will be required. Realistic quark masses must be used, and the discretization errors that result from calculating continuous functions with the discrete bits of a computer processor must be carefully assessed. The dependence of the force on spin, which the authors averaged over, and velocity, which is known empirically to be quite complicated, must also be calculated.

Once all that is achieved, an exciting applica-

tion will be to calculate the forces that occur in more extreme conditions than can be studied in terrestrial laboratories, such as those involving unstable hyperons. These particles are analogues of protons and neutrons that contain a valence strange ( $s$ ) quark instead of a  $u$  or  $d$ , and are very difficult to study experimentally. Hyperons are copiously produced during supernova explosions, and they occur inside neutron stars, where the enormous ambient pressure stabilizes them. Our ignorance of hyperon behaviour renders current models of supernovae and neutron stars wildly uncertain. Numerical QCD could sharpen our insight into these and many other issues in nuclear astrophysics.

It will also be fun, and important, to consider how forces get modified as we change the masses of the  $u$ ,  $d$  and  $s$  quarks, apply pressure, or otherwise tinker with the underlying theory. Over what range of parameters does the hard core persist, and so allow complex nuclear chemistry? That will provide fresh ideas in our understanding of how 'fine-tuned' the fundamental parameters must be in order to produce a user-friendly Universe. Such investigations might also provide something that is still conspicuously missing, despite the recent dramatic progress<sup>1</sup> — a simple qualitative explanation for the existence of the hard core.

Frank Wilczek is in the Department of Physics, Massachusetts Institute of Technology, 5 Cambridge Center, Cambridge, Massachusetts 02139, USA.  
e-mail: wilczek@mit.edu

1. Ishii, N., Aoki, S. & Hatsuda, T. preprint available at [www.arxiv.org/nucl-th/0611096](http://www.arxiv.org/nucl-th/0611096) (2006).
2. Yukawa, H. *Proc. Phys.-Math. Soc. Jpn* **17**, 48 (1935).

## DEVELOPMENTAL BIOLOGY

### Marked from the start

Not all cells in the early mammalian embryo are created equal. Even at the four-cell stage, embryonic cells that follow a particular pattern of division already have their developmental fate assigned to them. No cell will contribute exclusively to a specific cell type in the later embryo. But the progeny of some cells make a greater contribution to the 'inner cell mass' — the stem cells destined to become the fetus — and its surrounding 'trophoblast', which forms extraembryonic structures such as the placenta. The progeny of other cells will make a greater contribution to other extraembryonic structures.

Reporting on page 214 of this issue, Maria-Elena Torres-Padilla and colleagues find that the key to the cells' destiny lies, at least in part, outside their genes

(M.-E. Torres-Padilla *et al.* *Nature* **445**, 214–218; 2007).

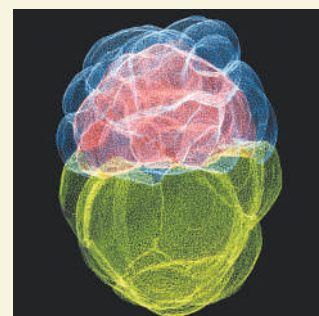
To fit into the nucleus, DNA is wound around histone proteins. Both the DNA and the histones can be studded with a variety of additional chemical groups — notably, methyl groups — that affect how tightly the structure is packed. These 'epigenetic marks' determine how accessible the genes in certain regions are, and they can interact directly with gene regulatory factors to activate or silence nearby genes.

Specific patterns of these marks are associated with particular cell fates in later stages of development, and with allowing stem cells to maintain the ability to develop into many different cell types (pluripotency). So Torres-Padilla *et al.* speculated that epigenetic instructions might

also help to determine the fate of early embryonic cells.

The authors concentrated on an epigenetic mark related to gene activation — methylation of an arginine amino acid in the histone H3 protein. They looked for differences between the cells destined for different fates in mouse embryos. An embryo is pictured here at the 32-cell stage, with the inner cell mass shown in red and the two types of trophoblast, polar and mural, shown in blue and green respectively.

In the four-cell embryo, methylation of H3 was highest in cells that were due to become the inner cell mass and polar trophoblast, and lowest in cells destined to contribute to the mural trophoblast. To see whether this epigenetic mark really affected



developmental fate, the authors manipulated the cells to overexpress the enzyme that carries out arginine methylation in two-cell embryos. This caused all the cells' progeny to become part of the inner cell mass — and increased the expression of certain proteins associated with pluripotency.

Epigenetic marks thus seem to be among the first developmental instructions that the embryonic cell receives.

Helen Dell





## 50 YEARS AGO

An almost unexplored record of the Earth's history is preserved in the sediment under the great oceans. Profiles of the uppermost layers have been obtained by many different types of corer, and to this list Prof. B. Kullenberg has now added a modified version of his elegant piston corer which should reach deeper than ever before... The new corer carries the same weight as the old model, 1,500 kgm., but has greater penetrating power, because with a length of only 2.3 metres there is a relatively small area of wall in frictional contact with the sediment... The corer sinks down through the sediment to a depth determined by the length of cable wrapped around its upper end. When all this cable has been unwound, the piston, which up to now has been locked at the lower end of the corer, is released and a core is taken. Although this ingenious corer has reached down to 29 m., and is expected to reach 50 m. in soft sediment... there still remains the ultimate problem of how to reach and take a continuous profile of sediment which can be a full kilometre thick. From *Nature* 12 January 1957.

## 100 YEARS AGO

The issue of *Science* for November 23 contains an article by Prof. McKeen Cattell on the selection, and arrangement in order of merit, of a thousand American men of science... Prof. Cattell also investigates the geographical distribution of American men of science according to place of birth... The production or "birth rate" of men of science per million of the population ranges from about 109 in Massachusetts—which stands far above the other States—and eighty-seven in Connecticut down to rates of only one or two in several of the southern States. It is argued that differences in stock can scarcely be great enough to account for this, and that accordingly the production of scientific men must be largely a matter of circumstance. From *Nature* 10 January 1907.

## EVOLUTIONARY BIOLOGY

# Genetics and bisexuality

Vincent Savolainen and Laurent Lehmann

**A population-genetic model indicates that if there is a gene responsible for homosexual behaviour it can readily spread in populations. The model also predicts widespread bisexuality in humans.**

For human societies at large, homosexuality is a sensitive issue. For biologists it is an intriguing one<sup>1,2</sup>. How can genes influencing homosexual — and so non-reproductive — behaviour be favoured by natural selection? An answer is offered by Gavrillets and Rice in a paper that has just appeared in *Proceedings of the Royal Society*<sup>3</sup>. They provide a population-genetic analysis that explains why, in theory, a gene predisposing an individual to homosexual behaviour would spread in a population, and that predicts its widespread occurrence in humans and other sexually reproducing species.

No predisposing gene for homosexual behaviour has been identified, but there is evidence that genetic controls are involved: for example, human twins are more likely both to be gay compared with non-identical brothers; and male homosexuality is more often inherited maternally, indicating that heritable maternal effects and/or genes linked to the X chromosome are in operation<sup>2,3</sup>. However, unlike heterosexuals, who devote a significant amount of time to reproductive sex, homosexuals are involved in non-reproductive sex, hampering the direct transmission of any gene underlying this behaviour. Homosexuality has a cost to fitness — that is, the ability of an individual to produce offspring that survive and reproduce — and it can only evolve if it otherwise provides indirect benefits to reproduction.

Three main mechanisms have been proposed in which variety in genes controlling homosexuality could be maintained in a population: overdominance, sexually antagonistic selection, and kin altruism<sup>2-4</sup>. For simplification, we will consider here male homosexuality, but these mechanisms also apply to female homosexuality. They also apply no matter how many genes contribute, but Gavrillets and Rice's analysis deals with a single theoretical gene and its two variants (alleles).

First, in the case of overdominance, a 'gay allele' would result in homosexual behaviour in an individual who has received this allele from both parents (homozygous), but would provide an advantage to the heterozygote (where only one parent has transmitted the gay allele). This situation would be similar to the renowned example of sickle-cell anaemia in Africa, a genetically inherited disease controlled by a deficient allele. Homozygotes for this allele suffer severe disorders. But because this allele confers resistance to malaria when heterozygous, it is maintained in human populations exposed to malaria. Under this scenario, heterozygotes

for the gay allele may have higher success in attracting females and/or their sperm may have some competitive advantage<sup>5</sup>.

In the second case, sexually antagonistic selection, a gay allele would result in a cost when expressed in males ('feminization' and loss of fitness), which would be counterbalanced by a fitness advantage when it is expressed in females.

In the third hypothesis, kin altruism, homosexuals would help their own family members, increasing the fitness of their relatives and therefore the probability that a gay allele is passed on to the next generation<sup>2,4</sup>.

These hypotheses have previously been speculative, but they have now been modelled and formalized by Gavrillets and Rice<sup>3</sup>. The authors adapted the classical population-genetic equations established by J. B. S. Haldane<sup>6,7</sup> and describe the evolution of the frequency of two alleles at one locus, in large populations for which each allele may result in sex-specific effects on fitness. Considering hypothetical straight and gay alleles, Gavrillets and Rice document the conditions of relative costs and benefits to fitness under which the gay allele can enter a population of straight alleles and be maintained subsequently. They establish the conditions under both the overdominance and sexually antagonistic-selection hypotheses for a homosexual gene that would be located on autosomes (non-sexual chromosomes) or on the X chromosome. These conditions still remain to be evaluated in the kin-altruism hypothesis.

Crucially, in these population-genetic models, a gay allele will produce variable degrees of homosexual behaviour, which is equivalent to the fitness cost of that behaviour (which, for example, could be interpreted as the proportion of time devoted to homosexual rather than reproductive sex). If one homozygous individual is not at all involved in reproductive sex, then the cost of homosexuality is maximal and this individual's phenotype is obviously strictly gay; however, in all other combinations, homozygous individuals exhibit a degree of bisexual behaviour depending on the costs.

Gavrillets and Rice show that, for a large set of costs and benefits, the gay allele can invade a population. Under overdominance, once a gay allele has entered a population it will be maintained in a polymorphic equilibrium, and this is easier if the homosexual gene is autosomal rather than X-linked. Further, under sexually antagonistic selection, the gay allele may even

go to fixation — that is, each individual will become homozygous for this allele — thus implying widespread bisexuality.

This theoretical framework<sup>3</sup> is an advance in evolutionary biology and studies of human behaviour because it generates several testable predictions: for example, if a gene influencing homosexuality is linked to the X chromosome, then it would support the sexual-antagonism hypothesis rather than overdominance. The framework will be used to guide research on the genetic basis of male and female homosexuality, and will help in resolving the 'Darwinian

paradox of male homosexuality'<sup>2</sup>. But it is of course theory only. Tasks for the future are to establish more precisely the costs and benefits of such behaviour in natural populations<sup>1</sup>. Such knowledge will help fine-tune these models of sexual orientation and show whether overdominance or antagonistic selection has been operating in mammals and throughout human history.

Vincent Savolainen is at the Jodrell Laboratory, Royal Botanic Gardens, Kew, Richmond TW9 3DS, UK. Laurent Lehmann is in the Department of Genetics, University of

Cambridge, Cambridge CB2 3HE, UK.

e-mails: v.savolainen@kew.org;

l.lehmann@gen.cam.ac.uk

1. Bagemihl, B. *Biological Exuberance: Animal Homosexuality and Natural Diversity* (St Martin's Press, New York, 1999).
2. Camperio-Ciani, A., Corna, F. & Capiluppi, C. *Proc. R. Soc. Lond. B* **271**, 2217–2221 (2004).
3. Gavrillets, S. & Rice, W. R. *Proc. R. Soc. Lond. B* **273**, 3031–3038 (2006).
4. Wilson, E. O. *Sociobiology: The New Synthesis* (Harvard Univ. Press, 1975).
5. MacIntyre, F. & Estep, K. W. *Biosystems* **31**, 223–233 (1993).
6. Haldane, J. B. S. *Trans. Camb. Phil. Soc.* **23**, 19–41 (1924).
7. Karlin, S. *Am. Math. Mon.* **79**, 699–739 (1972).

## OCEANOGRAPHY

# A marine nitrogen cycle fix?

Douglas G. Capone and Angela N. Knapp

**Some of our suppositions about the marine nitrogen cycle may be wrong. An innovative analysis of nutrients at the ocean's surface reveals a feedback mechanism that might hold the whole cycle in balance.**

The flow of nitrogen compounds between the oceans and the atmosphere is central to life, as nitrogen is a fundamental component of biomass and is essential for many biological processes. Although we have learned much about the nitrogen cycle of the oceans, two burning questions remain unanswered: is the marine nitrogen budget currently in balance, and are the processes that add and remove nitrogen to and from the seas closely linked? On page 163 of this issue, Deutsch *et al.*<sup>1</sup> provide evidence that the primary process responsible for putting nitrogen compounds into the sea — biological nitrogen fixation — is intimately associated, both geographically and temporally, with marine nitrogen removal. Furthermore, their work implies that the ratio of nitrogen to phosphorus in sea water may be the central factor that regulates nitrogen fixation.

Biological nitrogen fixation — the enzyme-catalysed reduction of nitrogen gas ( $N_2$ ) — continually adds nitrogen to the sea in the form of compounds that can be used as nutrients. Nitrogen fixation is commonly associated with certain cyanobacteria that inhabit the warm, sunlit surface waters of low-latitude oceans<sup>2</sup>. These photosynthetic bacteria can tap the immense reservoir of dissolved  $N_2$  gas in sea water, but their growth is often limited by the scarcity of other nutrients such as phosphorus and iron<sup>2</sup>.

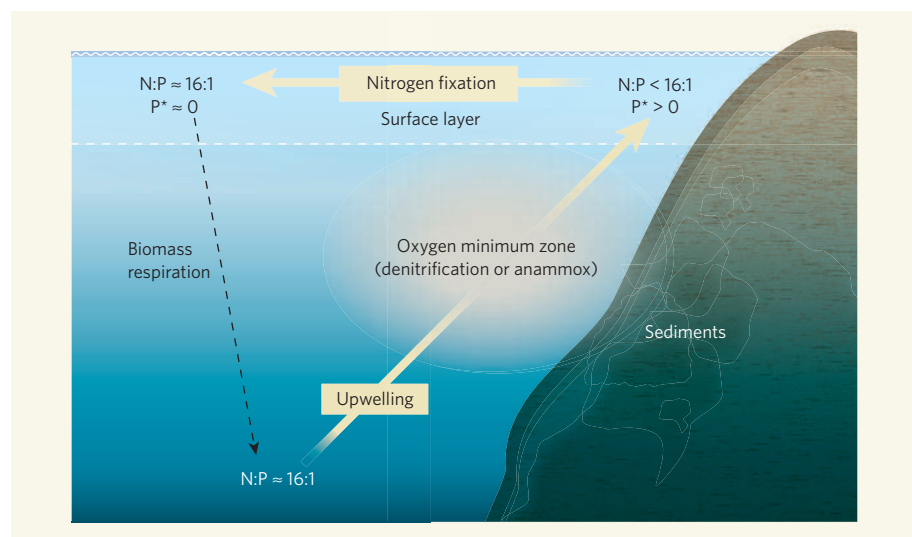
Fixed nitrogen is eventually converted to nitrate by nitrifying bacteria. Ultimately, the loss of nitrogen from the ocean occurs by denitrification, a process that converts nitrogen compounds such as nitrate back to  $N_2$ . Denitrification occurs mostly at depths of 200–700 metres in the 'oxygen minimum zones' (OMZs) of the ocean — that is, in the eastern tropical north Pacific, the eastern

tropical south Pacific and the Arabian Sea — and in marine sediments<sup>3–5</sup>.

Nitrogen fixation and denitrification are generally assumed to dominate the flows of nitrogen into and out of the ocean<sup>3,4,6</sup>. Some estimates suggest that the rate of marine denitrification substantially exceeds that of nitrogen fixation<sup>3,4</sup>. However, if this is the case, the ocean

would be progressively depleted of biologically available nitrogen, which is not supported by geochemical evidence<sup>5</sup>. If we accept that, in the oceanic nitrogen cycle, sources need to balance sinks<sup>5,6</sup>, then either the inputs are underestimated or the outputs are overestimated.

For the inorganic compounds dissolved in deep ocean water, the ratio of nitrogen to phosphorus is expected to be about 16:1, as noted by the oceanographer Alfred Redfield<sup>7</sup> in the early 1930s. Several of Deutsch's co-authors have previously analysed inorganic nutrients in oceanic subsurface waters, using a parameter  $N^*$  to discern areas with nitrogen-to-phosphorus ratios higher or lower than the 'Redfield' ratio<sup>8</sup>. They discovered large regions that have elevated nitrogen-to-phosphorus ratios, particularly in the north Atlantic, and speculated that these nutrient patterns resulted from the subsurface



**Figure 1 | Coupling between marine nitrogen fixation and denitrification zones.** Nitrogen enters the marine nitrogen cycle by 'fixation' when microbes at the ocean's surface convert nitrogen gas ( $N_2$ ) into biomass. This biomass may then release nitrogen compounds into the ocean. Denitrification and 'anammox' processes remove nitrogen from the sea by converting inorganic nitrogen compounds into  $N_2$ ; this occurs in the 'oxygen minimum zones' (OMZs) of deeper waters and in marine sediments. The expected ratio of nitrogen to phosphorus in the deep ocean is about 16:1. Deutsch *et al.*<sup>1</sup> define a parameter  $P^*$  that identifies deviations of relative phosphorus levels from this ratio. In upwelling waters in contact with OMZs, the nitrogen-to-phosphorus ratio is less than 16:1, and surface waters over the OMZs have an excess  $P^*$  ( $P^* > 0$ ). As these surface waters flow offshore, the 'excess' of phosphorus decreases back to the expected ratio ( $P^* \approx 0$ ), presumably as nitrogen fixers extract phosphorus without absorbing any nitrogen compounds. Residual nitrogen and phosphorus compounds are used by phytoplankton, which redeliver some of these compounds, with a nitrogen-to-phosphorus ratio of about 16:1, to the deep sea.



degradation of nitrogen-fixer biomass carried down from overlying waters. Such biomass typically has a nitrogen-to-phosphorus ratio greater than 16:1 (ref. 9). These authors<sup>8</sup> also proposed that the geochemically inferred high rates of nitrogen fixation in the north Atlantic were a result of unusually high growth rates of nitrogen-fixers, stimulated by the high flux of iron-bearing dust that is blown into this basin from northwest Africa. Although nitrogen fixation rates based on the value of  $N^*$  were estimated for the north Atlantic, nitrogen fixation rates in the Pacific were obscured by higher net rates of denitrification<sup>1,8</sup>.

On the basis of field observations of nitrogen fixation<sup>2,10</sup> and these  $N^*$  distributions<sup>2,8</sup>, the dominant sites of nitrogen fixation and denitrification seemed to be geographically removed from each other and were presumably coupled only over the timescales of ocean circulation (thousands of years). But recent modelling efforts have led to the conclusion that nitrogen fixation and denitrification must be more closely linked than this, and that nitrogen fixation provides a strong negative feedback that stabilizes the marine nitrogen cycle<sup>5,6</sup>, as originally hypothesized by Redfield<sup>7</sup>.

Deutsch *et al.*<sup>1</sup> now define a parameter  $P^*$  that discerns excesses or deficiencies of phosphorus relative to the canonical 16:1 ratio of nitrogen to phosphorus in sea water, and use this to examine nutrient fields at the ocean's surface. They note that, downstream of OMZs, surface waters that initially carry a surplus of phosphorus because of subsurface denitrification gradually lose this excess (Fig. 1). They attribute this effect to nitrogen fixation restoring the system to a 'Redfieldian' balance.

Deutsch and colleagues' analysis creates a very different view of the distribution of nitrogen fixation in the ocean and of the extent of its coupling to denitrification. The highest rates of fixation are not in the tropical Atlantic with its large inputs of iron-rich dust, as was previously surmised, but downstream from OMZs in the Pacific Ocean — despite the lower supply of airborne iron to these regions. So a supply of iron-rich dust may not exert as much influence on marine nitrogen fixation as is currently assumed<sup>11</sup>. These findings also indicate that, on a global scale, nitrogen fixation should compensate for denitrification, so balancing the marine-nitrogen books<sup>5,6</sup>.

Deutsch and colleagues' proposed distribution of nitrogen fixation<sup>1</sup> is supported by other evidence. For example, the distribution of nitrogen isotopes<sup>12</sup> — and, in some cases, that of oxygen isotopes<sup>13</sup> — in nitrate samples collected above the eastern tropical north Pacific and Arabian Sea OMZs is consistent with nitrogen-fixation inputs to these regions. Furthermore, the occurrence of dense surface accumulations of the nitrogen-fixing *Trichodesmium* cyanobacteria<sup>14</sup>, mapped by satellite, is strikingly similar to the  $P^*$  distributions presented by Deutsch *et al.*<sup>1</sup>.

As with any method, there are limits to the

uses of  $P^*$ . The authors note that  $P^*$  would underestimate nitrogen fixation by microbes that have both an unusually high nitrogen-to-phosphorus ratio and a very high export efficiency — that is, the capacity to sink rapidly into deep waters. Nitrogen-to-phosphorus ratios are highly variable in *Trichodesmium*<sup>9</sup>, but are unknown for other marine nitrogen-fixers. With regard to export efficiency, in some regions nitrogen fixation can occur in symbiotic groups of cyanobacteria and diatom algae<sup>10</sup>; these associations have a greater potential for gravitational sinking and export than free-living nitrogen fixers. Furthermore, surface  $P^*$  distributions do not take into account nitrogen fixation that may occur in environments that are uncoupled from the ocean's surface, such as deep-sea-floor methane seeps and hydrothermal vents<sup>15</sup>.

This re-evaluation of nitrogen fixation<sup>1</sup> arrives at the same time as other bombshells about the marine nitrogen cycle, including recognition that nitrogen-fixing microbes are more diverse than was thought<sup>7</sup>, and the apparent importance in several OMZ studies<sup>16</sup> of the anaerobic ammonia oxidation or 'anammox' pathway for nitrogen removal (as opposed to conventional denitrification). Recent evidence also shows that Archaea microbes, which are distinct from bacteria, also have leading roles in the marine nitrogen cycle, as they predominate in nitrification<sup>17</sup> and may be involved in nitrogen fixation<sup>15</sup>. All things considered, it may soon be time to start rewriting the textbooks. ■

Douglas G. Capone and Angela N. Knapp are in the Department of Biological Sciences and at the Wrigley Institute for Environmental Studies, University of Southern California, Los Angeles, California 90089, USA. Douglas G. Capone is currently on sabbatical at the Laboratoire d'Océanographie de Villefranche, France. e-mails: capone@usc.edu; angelakn@usc.edu

- Deutsch, C., Sarmiento, J. L., Sigman, D. M., Gruber, N. & Dunne, J. P. *Nature* **445**, 163–167 (2007).
- Mahaffey, C., Michaels, A. & Capone, D. G. *Am. J. Sci.* **305**, 546–595 (2005).
- Codispoti, L. *et al. Sci. Mar.* **65**, 85–105 (2001).
- Brandes, J. A. & Devol, A. H. *Glob. Biogeochem. Cycles* **16**, doi:10.1029/2001GB001856 (2002).
- Deutsch, C. *et al. Glob. Biogeochem. Cycles* **18**, doi:10.1029/2003GB002189 (2004).
- Gruber, N. in *The Ocean Carbon Cycle and Climate* (eds Follows, M. & Oguz, T.) 97–148 (Kluwer Academic, Dordrecht, 2004).
- Redfield, A. C. in *James Johnstone Memorial Volume* (ed. Daniel, R. J.) 176–192 (Liverpool Univ. Press, 1934).
- Gruber, N. & Sarmiento, J. L. *Glob. Biogeochem. Cycles* **11**, 235–266 (1997).
- Krauk, J., Sohm, J., Montoya, J., Villareal, T. & Capone, D. G. *Aquat. Microbial Ecol.* **42**, 243–253 (2006).
- Carpenter, E. J. *et al. Mar. Ecol. Prog. Ser.* **188**, 273–283 (1999).
- Berman-Frank, I., Cullen, J., Harel, Y., Sherrell, R. & Falkowski, P. G. *Limnol. Oceanogr.* **46**, 1249–1277 (2001).
- Brandes, J., Devol, A., Yoshinari, T., Jayakumar, D. & Naqvi, S. *Limnol. Oceanogr.* **43**, 1680–1689 (1998).
- Sigman, D. M. *et al. Glob. Biogeochem. Cycles* **19**, GB4022 (2005).
- Westberry, T. K. & Siegel, D. A. *Glob. Biogeochem. Cycles* **20**, GB4016 (2006).
- Mehta, M. P. & Baross, J. A. *Science* **314**, 1783–1786 (2006).
- Dalsgaard, T., Thamdrup, B. & Canfield, D. *Res. Microbiol.* **156**, 457–464 (2005).
- Wuchter, C. *et al. Proc. Natl Acad. Sci. USA* **103**, 12317–12322 (2006).

## BIOINFORMATICS

# Industrializing neuroscience

Henry Markram

**The project for producing a genome-wide atlas of gene expression in the mouse brain shows how, with advancing technology, huge volumes of data can be collected and made accessible through the Internet.**

It took until the beginning of the twenty-first century for the Industrial Revolution to reach life science, an event marked by the sequencing of the human genome<sup>1,2</sup>. The wave of industrialization is sweeping on, not least in neuroscience, where it is especially evident in the Allen Institute for Brain Science's effort to map the 'transcriptome' of the entire mouse brain — the brain locations where each of 21,500 genes is activated. Following online publication last month, results from the project appear on page 168 of this issue<sup>3</sup>.

The project essentially involved building 21,500 three-dimensional brain atlases, plus a reference atlas — a task that would probably take a single researcher a century to complete. The Allen Institute<sup>4</sup>, however, decided to complete this task in a couple of years by industrializing the work flow. The brains of mice were

sectioned into 25- $\mu$ m-thick slices and stained for an activated gene using a technique called *in situ* hybridization<sup>5</sup>. To obtain the three-dimensional locations of all 21,500 activated genes, more than 6,000 brains were sectioned and processed — a total of 1 million sections were finally used.

The sections were fed into a pipeline for automated staining that processes more than 16,000 sections per week. Every micrometre of each section was then photographed at low and high magnifications using automated microscopy and image capture. Operating continuously, this system produced a total of 85 million images, which were dumped onto storage servers at a rate of a terabyte (a trillion bytes) of data per day during full production, to end with more than 600 terabytes of data. The next pipeline screened the quality of the data to

provide approved data for final viewing through an elaborate website<sup>6</sup>.

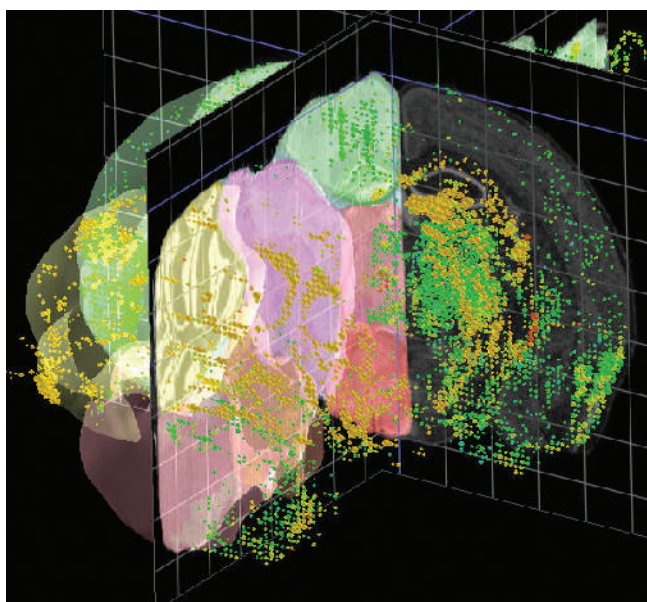
The resulting Allen Brain Atlas (ABA) allows one to view all the locations at which any of the 21,500 genes in the mouse brain are activated, down to cellular-level resolution. The site allows one to search any anatomical region of the brain for any gene or combination of genes. The activated genes can be viewed with tools that can zoom from a whole-brain section down to a single cell while retrieving data at multiple levels of resolution — itself a masterly stroke of informatics. The site is also linked to many gene banks around the world, allowing researchers to zoom into the building-blocks of each gene.

The ABA team also constructed a three-dimensional brain viewer, which maps the location and intensity of gene expression onto a reference atlas. The viewer can manipulate the atlas in three dimensions while moving a slider to virtually slice through any location in the brain and focus on areas of interest (Fig. 1).

Overall, the ABA provides unprecedented insight into how gene expression is orchestrated throughout the brain. The study does not determine specifically how sensitive the method is, but it was sensitive enough to yield a major surprise: that most of the genome is activated in the building of the brain. Expression of nearly 80% of the genes was detected, much higher than expected<sup>7</sup>. This is unlikely to be the case for body cells, as they do not require neuron-specific proteins (such as those encoding synapses and ion channels, and those regulating neuronal plasticity). Individual brain cells, however, express a much smaller percentage of genes (about 15%), which still allows for an enormous molecular diversity of neurons, connections and brain regions. Theoretically, the potential molecular diversity is greater than  $10^{4,000}$ , which should indicate that such gene-expression atlases will only be properly understood in combination with expression rules that make up the biologically viable solution — the choices made in evolution to focus on a minute subset of possible expression combinations.

The real power of the ABA, therefore, lies in its ability to pinpoint the anatomical locations of expressed genes. Taken together, these expression brain atlases provide the recipe of genes that are most needed to construct each brain region. On the other hand, each individual expression atlas highlights the different brain regions where the gene is used and reveals that some fundamental genes are activated in almost all regions, whereas other more specialized genes are activated only in specific locations — an essential guide for future transgenic experiments.

When such expression atlases are combined



**Figure 1 | Three-dimensional gene expression in the mouse brain.** The model shows the expression of the gene encoding a G-protein-coupled receptor (GPCR12) mapped onto a reference atlas. On the left side are three-dimensional structures of different brain regions (in different colours). The expression of GPCR12 is indicated by the dots on both sides of the brain. The colour of the dots indicates the expression level — green (low expression) through to red (high expression). The squared planes can be moved along the *x* and *y* planes to focus on different parts of the brain.

with functional studies on the proteins produced, we may begin to understand the essential functions that each gene has in different brain regions, and understand why certain genes are not expressed in certain brain regions and in certain types of cell. By analysing the overall correlation between anatomical locations for all 21,500 genes, it may be possible to get a glimpse of the role played by the ‘conductor’ that orchestrates gene expression throughout the brain. If such atlases could be built for each step taken in development, through adulthood and into ageing, we could capture the entire ‘symphony’ as genes are switched

on and off in different parts of the brain during the life of an animal, and begin to understand the expression rules.

The ABA’s main shortcoming is that any given cell in the brain is tested for only one expressed gene. The ideal case would be to determine the expression of all 21,500 genes for each neuron in the mouse brain — a task that is some 100-million-fold larger. With improved technology, such a feat may become possible in the future. The potential of such data would be incredible. By comparing how the transcriptomes of different types of neuron are correlated with their emergent properties, the secrets of how different neurons are built and connected might reveal themselves at the genetic level.

The ABA marks the beginning of an era in neuroscience in which such tasks become possible. We will see the production of hitherto unimaginable volumes of precisely standardized data, which will allow bottom-up reconstruction and eventual modelling of the brain in

exquisite biological detail.

Henry Markram is at the Brain Mind Institute, Ecole Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland.  
e-mail: henry.markram@epfl.ch

1. International Human Genome Sequencing Consortium *Nature* **409**, 860–921 (2001).
2. Venter, J. C. *et al.* *Science* **291**, 1304–1351 (2001).
3. Lein, E. S. *et al.* *Nature* **445**, 168–176 (2007).
4. [www.alleninstitute.org](http://www.alleninstitute.org)
5. Visel, A., Thaller, C. & Eichele, G. *Nucleic Acids Res.* **32**, D552–D556 (2004).
6. [www.brain-map.org](http://www.brain-map.org)
7. Sandberg, R. *et al.* *Proc. Natl Acad. Sci. USA* **97**, 11038–11043 (2000).

## MATERIALS SCIENCE

# Displaced by radiation

Rodney C. Ewing

**The mineral zircon suffers more structural damage from the  $\alpha$ -decay of plutonium present in its crystal than was thought. That could have a knock-on effect on strategies for managing nuclear waste.**

On page 190 of this issue, Farnan, Cho and Weber<sup>1</sup> describe how nuclear magnetic resonance (NMR) spectroscopy can be used to assess the damage caused to a solid’s structure by  $\alpha$ -decays of an emitter incorporated into its crystal. The authors test their technique on the mineral zircon ( $\text{ZrSiO}_4$ ), and find that each  $\alpha$ -decay event displaces significantly more atoms than simulations had predicted.

Zircon is an important material for two reasons. First, its trace content of long-lived  $\alpha$ -emitters, combined with its ubiquity and high chemical durability in Earth’s crust, has made it the mineral most often used in geological dating. Second — and this lends Farnan and colleagues’ work its most immediate significance — zircon has been proposed as a material in which to immobilize the plutonium isotope



$^{239}\text{Pu}$  over geological timescales. This  $\alpha$ -emitter, of half-life 24,110 years, is the environmentally and politically awkward waste product of uranium-fuelled nuclear reactors. An accurate assessment of how its decay — as well as those of other members of the group of elements known as actinides<sup>2</sup>, among them neptunium, curium and americium — might compromise the long-term chemical durability of surrounding materials is essential.

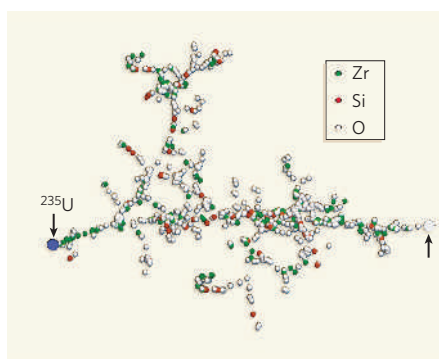
Such  $\alpha$ -decays can lead to the slow accumulation of 'damage domains' and the eventual complete loss of periodic structure in a solid crystal, causing changes in the material's physical and chemical properties<sup>3–6</sup>. As long ago as 1815, the Swedish chemist Jöns Berzelius noted a rapid, and at the time inexplicable, release of heat from minerals that had accumulated such radiation damage. In 1893, the Norwegian mineralogist Waldemar Brögger coined the term 'metamict' to describe amorphous materials that had, judging by their well-formed crystal faces, clearly at some point been crystalline.

With the construction of the first nuclear reactors during the Second World War, larger-scale radiation effects were seen in solids. The intense neutron flux from the production reactors at Hanford, Washington state, which were used to produce fissile material for the first nuclear bombs, displaced atoms in the surrounding graphite moderator and reactor fuel, causing both to swell. This effect was termed 'Wigner's disease', after Eugene Wigner, the theoretical physicist who first predicted these radiation effects.

The damage from an  $\alpha$ -decay event comes from two different particles. First, there is the  $\alpha$ -particle itself. This has a relatively low mass — it is a helium nucleus of two neutrons and two protons — and dissipates most of its energy (around 5 megaelectronvolts, MeV) through electronic excitations along its path, which is just a few tens of micrometres long. Only near the end of its trajectory does it interact with the crystal atoms in billiard-ball fashion, causing a few hundred displacements.

Second, there is the 'recoil nucleus' that is left after the emission of the  $\alpha$ -particle, for example uranium-235 following the decay of plutonium-239. This particle has a much lower energy, just 86 keV, but deposits nearly all of it through ballistic interactions along a path of just a few tens of nanometres. This "canon ball awry through a crowded dance floor" — to use chemistry Nobel laureate Roald Hoffmann's metaphor<sup>7</sup> — can cause several thousand atomic displacements. Energy deposition in this recoil cascade is high, at a few electronvolts per atom, and temperatures in its core can exceed 5,000 K for some picoseconds (a picosecond is  $10^{-12}$  s).

The damage comprises a complex combination of isolated defects caused by the  $\alpha$ -particle and branching collision cascades from the recoil nucleus (Fig. 1). The exact extent of the damage depends on the temperature of the irradiation. As they accumulate, the defects



**Figure 1 | Defect trail.** The calculated spatial distribution of vacancies (colour-coded for zirconium, silicon and oxygen) produced by the recoil of a typical 86-keV uranium-235 nucleus in zircon following the  $\alpha$ -decay of plutonium-239 (ref. 8). The arrow at the extreme right indicates the original position of the recoil nucleus. This simulation is based on a two-body collision approximation, which provides a good representation of the spatial extent of the cascade and the branching through the formation of sub-cascades. Farnan and colleagues' NMR experiments indicate that even more atoms are displaced along such recoil tracks than was thought — leading to the swifter formation of amorphous domains in the host material. (Figure courtesy of W. J. Weber.)

and collision cascades interact and finally overlap, producing an amorphous material. This is accompanied by significant changes in the material's physical properties: zircon's density, for example, decreases by some 17% at a dose of  $10^{19}$   $\alpha$ -decay events per gram (ref. 8).

Farnan and colleagues' work<sup>1</sup> addresses two limitations that have plagued studies of radiation damage in zircon. First, there is the extent to which the experiments performed up to now produce results that are comparable to effects seen in natural zircon. Natural zircon contains the  $\alpha$ -emitting actinides uranium and thorium, and has thus accumulated its radiation doses over many hundreds of millions of years. Synthetic phases doped with actinides such as the plutonium isotope  $^{238}\text{Pu}$  (half-life 87.7 years), on the other hand, require just 5–10 years for irradiation, and materials irradiated with ion beams reach radiation levels of interest in less than an hour. But can these accelerated actinide-doping and ion-beam methods be used to simulate the long-term effects important for evaluating nuclear-waste containment strategies over hundreds of thousands of years? Does the rate at which the radioactive dose is administered really play no role?

The second problem that hampers studies of the evolution of damage structure is that, as the dose increases, so the length-scale of interest — and the appropriate tool to investigate it — changes. At low doses, techniques such as Raman spectroscopy are required that are sensitive to small-scale structural distortions in the crystal. At intermediate doses, methods such as X-ray diffraction measure the expansion of a crystal's unit cell as gaps accumulate between

its component atoms, and high-resolution transmission electron microscopy shows the overlap of recoil cascades. At the highest doses, when the material is fully amorphous, X-ray absorption spectroscopy must be used to investigate the geometry of nearest neighbours. None of these methods alone can resolve the different types of defects and damaged domains.

Farnan *et al.*<sup>1</sup> measure the NMR spectrum of the silicon-29 isotope present in zircon doped with the long-lived  $^{239}\text{Pu}$  and the shorter-lived  $^{238}\text{Pu}$  using 'magic-angle spinning'. This technique enhances the resolution of an NMR spectrum by spinning the sample under investigation at high speeds at a certain angle to the applied magnetic field. Spinning highly radioactive samples at more than 200,000 rotations per minute is no mean feat and required elaborate safety procedures, including the triple containment of all the samples.

The improved resolution possible with magic-angle spinning means that even at low doses, before damage domains have overlapped, the nature of the individual damage events can be discerned. Furthermore, the NMR signal can be used to deconvolute the spectrum into the crystalline and amorphous fractions. The result that emerges is that around 5,000 atoms are displaced in each  $\alpha$ -decay event. This is significantly more than the 1,000–2,000 predicted by standard simulations.

Comparison of these data from doped zircons with data from natural zircons<sup>1</sup>, which are damaged over hundreds of millions of years, shows that the dose rate has no significant effect, despite its differing by a factor of  $10^8$ . Thus, accelerated experiments using highly radioactive actinides can indeed be used to simulate long-term effects in forms of nuclear waste. That conclusion applies to the wide range of structure types and compositions that are currently under investigation to immobilize the actinide components of nuclear waste<sup>9</sup>.

The increased damage observed by Farnan *et al.*<sup>1</sup> in each  $\alpha$ -decay event means that the amorphous state will occur in radioactively doped zircon sooner, rather than later. Such information is vital in assessing the durability of such materials, and in informing the development of strategies for the safe encapsulation of actinides.

Rodney C. Ewing is in the Department of Geological Sciences, University of Michigan, Ann Arbor, Michigan 48109-1005, USA. e-mail: rodewing@umich.edu

1. Farnan, I., Cho, H. & Weber, W. J. *Nature* **445**, 190–193 (2007).
2. Ewing, R. C. *Proc. Natl Acad. Sci. USA* **96**, 3432–3439 (1999).
3. Weber, W. J. *et al. J. Mater. Res.* **13**, 1434–1484 (1998).
4. Robinson, M. T. *J. Nucl. Mater.* **216**, 1–28 (1994).
5. Hobbs, L. W. *et al. J. Nucl. Mater.* **216**, 291–321 (1994).
6. Ewing, R. C. *et al. Rev. Mineral. Geochem.* **39**, 319–361 (2000).
7. Hoffmann, R. in *The Metamict State: Poems by Roald Hoffmann* 101–102 (Univ. Central Florida Press, Orlando, 1987).
8. Ewing, R. C. *et al. Rev. Mineral. Geochem.* **53**, 387–425 (2003).
9. Ewing, R. C. *et al. J. Appl. Phys.* **95**, 5949–5971 (2004).

## QUANTUM INFORMATION

## Source of triggered entangled photon pairs?

Arising from: R. M. Stevenson *et al.* *Nature* **439**, 179–182 (2006)

The realization of an entangled photon source will be of great importance in quantum information — for example, for quantum key distribution and quantum computation — and Stevenson *et al.*<sup>1</sup> have described such a source. However, we show here that first, their source is not entangled; second, they use inappropriate entanglement indicators that rely on assumptions invalidated by their data; and third, their source has insignificant entanglement even after simulating subtraction of the significant quantity of background noise. We therefore find that the standard of proof required for a semiconductor source of triggered entangled photon pairs has not been met by Stevenson *et al.*<sup>1</sup>.

First, Stevenson *et al.*<sup>1</sup> produce pairs of photons by radiative decay of biexcitons in single quantum dots, and measure strong coherences between two polarization populations, a necessary but not sufficient condition for entanglement. Establishing the degree<sup>2–5</sup> or presence<sup>6</sup> of entanglement between two qubits is a solved problem: Stevenson *et al.*<sup>1</sup> apply none of these standard techniques. The entanglement of formation<sup>2</sup>,  $E_F$ ; concurrence<sup>3,4</sup>,  $C$ ; and tangle<sup>5</sup>,  $T$ , are directly related (for example,  $C = \sqrt{T}$ ), and are quantitative, measuring the degree of entanglement. Other methods, such as the Peres condition<sup>6</sup>, or fidelity with a maximally entangled state, are qualitative, indicating but not measuring entanglement. Any of these quantities can be calculated directly from two-photon density matrices obtained

by quantum state tomography<sup>7</sup>, as measured in Fig. 3 of ref. 1.

For example, using the tangle to quantify unambiguously the degree of entanglement ( $0 \leq T \leq 1$ ;  $T=0$  for unentangled states and  $T=1$  for maximally entangled states<sup>4</sup>), we find that  $T=0$  for every measured density matrix<sup>1</sup>. There is no entanglement. The large observed coherences in dots B and C in Fig. 3b,d of ref. 1 indicate only classical correlation. State purity provides an independent check. Quantified by the linear entropy ( $0 \leq S_L \leq 1$ ;  $S_L = 0$  for pure states and  $S_L = 1$  for maximally mixed states<sup>5</sup>), states with  $\% < S_L$  always have strictly zero entanglement<sup>8</sup>. The states in Fig. 3 of ref. 1 are highly mixed, with  $0.92 \leq S_L \leq 0.99$ , and so are unentangled. Similarly, all other metrics and indicators listed here find no entanglement.

Second, Fig. 2c of Stevenson *et al.*<sup>1</sup> measures the degree of correlation for quantum dots A and B. Entanglement is indicated by a constant value that is greater than 50% (ref. 9). The authors arbitrarily fit a sinusoid to dot A, and a constant value to dot B ( $22.2 \pm 2.8\%$ ). (There are no statistical grounds for asserting that dot B's correlation is independent of measurement basis<sup>1</sup>. Dot B is better fitted with a sinusoid with the same period as dot A, which ranges from 20% to 25%;  $\bar{\chi}^2 = 0.065$  compared with  $\bar{\chi}^2 = 0.13$ .) Irrespective of curve choice, these data provide no evidence for entanglement, as the mean for dot B is  $9.9\sigma$  less than 50%. The curve shape is a very poor indicator of entangle-

ment, because it varies greatly even between different unentangled sources: dots A and B are both unentangled, albeit with different classical correlations.

Stevenson *et al.*<sup>1</sup> also use an eigenvalue entanglement indicator. Only if both individual photons are unpolarized is entanglement indicated by the largest eigenvalue of the two-photon density matrix exceeding 0.5. The only data of Stevenson *et al.*<sup>1</sup> that provide a direct measure of the polarization of the photons emitted by the dots are the density matrices in their Fig. 3. These are a precise measure and are self-consistent: the same data used by Stevenson *et al.*<sup>1</sup> to gauge entanglement can also be used to check the photon polarization. For  $\rho_{3d}$ , we obtain the polarization of each photon in the pair by tracing out the other, finding that one is partially polarized with a degree of polarization<sup>10,11</sup> of  $4.5 \pm 1.9\%$  — the eigenvalue method is therefore invalid. (Degree of polarization is the length of the Stokes vector<sup>10</sup>,  $\wp = (1 - 4|\rho_i|)^{1/2}$ , in which  $|\rho_i|$  is the determinant of the single-polarization density matrix,  $\rho_i$  (ref. 11).) Note that the other photon has zero polarization,  $0 \pm 1.1\%$ , owing to the artificial normalization imposed by Stevenson *et al.*<sup>1</sup> (Table 1).

Based on the data in their Figs 2 and 3, Stevenson *et al.*<sup>1</sup> claim that their measurements indicate that dots with small exciton splitting emit entangled photons. As we have shown, this conclusion is not supported by their data or methods.

**Table 1 | Density matrix of Fig. 3d of ref. 1 and the underlying probabilities\***

Measure	$P$	$\Delta P$	Measure	$P$	$\Delta P$	Measure	$P$	$\Delta P$	Measure	$P$	$\Delta P$
VV	0.30470	0.00606	LV	0.25592	0.00737	DL	0.23479	0.00919	HD	0.24041	0.01065
VH	0.19530	0.00496	LH	0.24408	0.00562	DD	0.30285	0.00463	HR	0.23647	0.01235
HH	0.32605	0.01084	DH	0.25058	0.00459	LD	0.26214	0.00631	VR	0.27020	0.00773
HV	0.17395	0.01409	DV	0.24942	0.00651	VD	0.26054	0.00511	LR	0.30477	0.00769

\*R. M. Stevenson *et al.*, personal communication.

Measure refers to polarization analyser settings for each photon: V, vertical linear; L, left circular; D, diagonal linear; H, horizontal linear; R, right circular. Standard unbiased tomography uses complete measurement sets to ensure that the sum of the probabilities is unity — for example,  $P_{VV} + P_{VH} + P_{HV} + P_{HH} = 1$ . Repeating this in each measurement basis allows for compensation of measurement drift. This is not done by Stevenson *et al.*<sup>1</sup>, who instead use an artificial normalization that sets the sum of each of their first four probability pairs to a half — for example,  $P_{VV} + P_{VH} = 0.5$ . The effect of their extra constraint is to force one photon in each pair to be unpolarized. The resulting density matrix,  $\rho_{3d}$ , is accordingly biased:

$$\rho_{3d} = \begin{bmatrix} 0.3261 & -0.0096 & -0.0101 & 0.1038 \\ -0.0096 & 0.1739 & 0.0009 & 0.0101 \\ -0.0101 & 0.0009 & 0.1953 & 0.0105 \\ 0.1038 & 0.0101 & 0.0105 & 0.3047 \end{bmatrix} + i \begin{bmatrix} 0 & -0.0135 & 0.0166 & 0.0002 \\ 0.0135 & 0 & -0.0235 & -0.0166 \\ -0.0166 & 0.0235 & 0 & 0.0202 \\ -0.0002 & 0.0166 & -0.0202 & 0 \end{bmatrix}$$

**Methods.** To perform unbiased quantum state tomography and a full uncertainty analysis, we need count data for the probabilities,  $P_i$ , in Table 1. Stevenson *et al.*<sup>1</sup> use second-order correlations,  $g^{(2)}(\tau=0) = C/(N/n)$ , between photons 1 and 2, to obtain  $P = 1/2 g^{(2)} / (g^{(2)} + g^{(2)})$ , where A and B refer to orthogonal measurements on photon 2; C is the number of pairs detected in the same laser cycle;  $N/n$  is the average number of pairs in different laser cycles;  $n$  is the number of measured finite delay peaks; the uncertainty is  $\Delta g^{(2)} = g^{(2)}(C^{-1} + N^{-1})^{1/2}$ , assuming  $\Delta n = 0$  and poissonian uncertainties,  $\Delta C = \sqrt{C}$ ,  $\Delta N = \sqrt{N}$  (R. M. Stevenson *et al.*, personal communication). The count data, C,  $N$  or  $n$ , was not available, the figure of 1,000 coincidences given in ref. 1 was simply provided as a guideline (R. M. Stevenson *et al.*, personal communication). Accordingly, it is not clear how  $\Delta P_i$  were obtained: for  $N$  counts, poissonian uncertainties are  $\sqrt{N}$ , and cannot be calculated from a probability. In the absence of counts, we used the artificially normalized probabilities to reconstruct  $\rho_{3d}$  for all other density matrices, we estimated the values of the elements directly from Fig. 3 of ref. 1. Linear entropy and tangle are calculated directly from the reconstructed/estimated density matrices<sup>3–5,7</sup>. We estimated uncertainties from an ensemble of 5,000 density matrices generated by creating a new data set by sampling from a gaussian distribution centred on  $P_i$ , with standard deviation equal to  $\Delta P_i$ , and by applying maximum-likelihood tomography to each such data set. The tangle and degrees of polarization were calculated for each of the 5,000 matrices; uncertainties are the standard deviation of these quantities. To model background subtraction, 5,000 physical density matrices were obtained after subtracting  $0.49 I/4$ , where  $I$  is the  $4 \times 4$  identity matrix.



Third, unwanted background light degrades entanglement. Stevenson *et al.*<sup>1</sup> identify 49% (no error given) of photon pairs as background due to dark counts and emission from layers other than the dot. An improved source was simulated by subtracting the projected (but not directly measured) number of background counts; the resulting density matrices and data are not given. Accordingly, we modelled this by subtracting 49% unpolarized light from  $\rho_{3d}$ : the simulated source has insignificant tangle,  $T = 0.028 \pm 0.022$ . Naturally, removing unpolarized light serves to increase further the polarization of the partially polarized photon, to  $8.8 \pm 3.4\%$  — the eigenvalue method remains invalid. Appropriate background subtraction<sup>12</sup> may indicate the

potential of an entangled source, but such a source is not useful in quantum information. For example, the security of the Ekert protocol in quantum cryptography requires actual, as opposed to virtual, entanglement.

Alexei Gilchrist\*, Kevin J. Resch\*†, Andrew G. White\*

\*Department of Physics and Centre for Quantum Computer Technology, University of Queensland, Brisbane, Queensland 4072, Australia  
e-mail: andrew.white@physics.uq.edu.au

†Present address: Department of Physics and Institute for Quantum Computing, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada

1. Stevenson, R. M. *et al. Nature* **439**, 179–182 (2006).

2. Bennett, C. H., DiVincenzo, D. P., Smolin, J. A. & Wootters, W. K. *Phys. Rev. A* **54**, 3824–3851 (1996).
3. Wootters, W. K. *Phys. Rev. Lett.* **80**, 2245–2248 (1998).
4. Coffman, V., Kundu, J. & Wootters, W. K. *Phys. Rev. A* **61**, 052306 (2000).
5. White, A. G., James, D. F. V., Munro, W. J. & Kwiat, P. G. *Phys. Rev. A* **65**, 012301 (2001).
6. Peres, A. *Phys. Rev. Lett.* **77**, 1413–1415 (1996).
7. James, D. F. V., Kwiat, P. G., Munro, W. J. & White, A. G. *Phys. Rev. A* **64**, 052312 (2001).
8. Munro, W. J., James, D. F. V., White, A. G. & Kwiat, P. G. *Phys. Rev. A* **64**, 030302 (2001).
9. Young, R. J. *et al. N. J. Phys.* **8**, 29 (2006).
10. Stokes, G. G. *Trans. Camb. Phil. Soc.* **9**, 399–416 (1852).
11. Mandel, L. & Wolf, E. in *Optical Coherence and Quantum Optics* 353–354 (Cambridge University Press, Cambridge, 1995).
12. Vasanelli, A., Ferreira, R. & Bastard, G. *Phys. Rev. Lett.* **89**, 216804 (2002).

doi:10.1038/nature05546

## QUANTUM INFORMATION

# Stevenson *et al.* reply

Replying to: A. Gilchrist, K. J. Resch & A. G. White *Nature* **445**, doi:10.1038/nature05546 (2007)

We have reported triggered photon-pair emission from single quantum dots that is suggestive of polarization entanglement<sup>1</sup>. Gilchrist *et al.*<sup>2</sup> criticize our analysis of the density matrix, claiming that the entanglement test is inappropriate. However, we show here that this analysis is in fact valid for an unpolarized source such as ours, and that this analysis (as well as alternative measures of entanglement) suggests that quantum-dot sources do emit entangled photons, and that their quality is rapidly improving<sup>3</sup>.

Gilchrist *et al.*<sup>2</sup> disregard the fact that direct measurement of the dot emission shows the source to be unpolarized, within error (Fig. 1). Measuring linearly polarized exciton and biexciton intensities as a function of rotation of a half-wave plate yields polarizations of  $0 \pm 0.5\%$  and  $0 \pm 1.1\%$ , respectively. Our use of the eigenvalue test of entanglement is therefore valid. A less precise, although consistent, polarization measure is determined from the photon-pair intensities shown by the two-photon density matrix in Fig. 3d of ref. 1. From this density matrix, we determine polarizations of  $0 \pm 4.6\%$  and  $4.5 \pm 4.6\%$  for the two photons, confirming independently that emission is unpolarized.

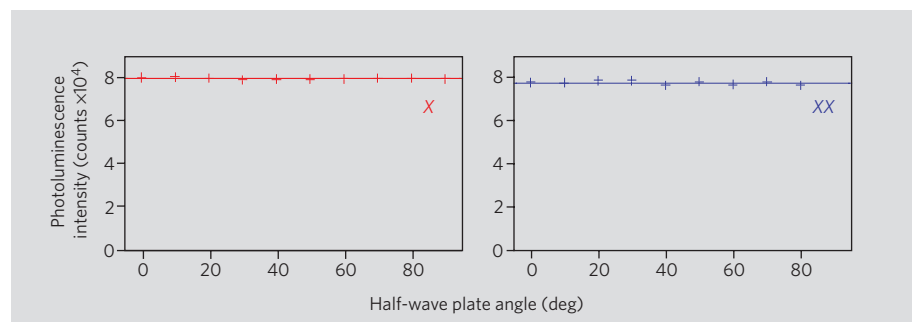
The eigenvalue is an intuitive test for entanglement and is simply the probability that the source emits into a single state, which naturally cannot exceed 50% for a classical source that has at least one unpolarized photon. We analysed light originating from the dot alone by subtracting contributions from other layers in the sample determined directly from the emission spectrum of the source<sup>1</sup>. Background correction is a well established procedure<sup>4</sup> and it is reasonable to assume that dot emissions can be isolated better with improved design

of devices<sup>3</sup>. We pointed out<sup>1</sup> that only after correction did the measured eigenvalue of  $0.58 \pm 0.04$  violate the 0.5 limit for an unpolarized classical source. The error on the background level of  $<1\%$  makes only a minor contribution to this error in the eigenvalue. For an unpolarized, partially uncorrelated classical source<sup>1</sup>, the maximum possible eigenvalue is only 0.4.

Using data supplied by us, Gilchrist *et al.*<sup>2</sup> calculate tangle  $T = 0.028 \pm 0.022$  and concur-

rence  $C = \sqrt{T} = 0.167 \pm 0.090$ . The values of  $T > 0$  and  $C > 0$  show that emission from the dot must contain entangled photons<sup>2</sup>, in general agreement with our results<sup>1</sup>.

Although we show that photon pairs originating from a quantum dot violate classical limits, our most important message<sup>1</sup> is the idea of how to manipulate quantum dots to generate entangled photons. By selecting dots with specific emission energy, or applying an appropriate magnetic field, the exciton polariza-



**Figure 1 | Measurements showing that emission is unpolarized, within error.** Linearly polarized exciton (X) and biexciton (XX) intensities were measured as a function of the angle of a half-wave plate. Intensities are independent of the wave-plate angle and thus of the linear polarization detection basis. The degree of polarization is therefore zero, with errors of 0.5 for the exciton photon and 1.1% for the biexciton photon. Note that the biexciton error is higher owing to greater sensitivity to excitation power fluctuations. A similar result is obtained for circular polarization. The probability of detecting polarized photon pair combinations was measured experimentally from pairs of second-order correlations with orthogonal exciton polarizations as described in ref. 1; that is,  $P = 0.5 g_{xx,xx}^{(2)} / (g_{xx,xx}^{(2)} + g_{xx,xx}^{(2)})$ . Corresponding errors are determined by standard analysis of the correlation function errors. Because the emission is unpolarized, this normalization compensates for excitation drift and allows formation of a density matrix for the emission; 16 such probability measurements were recorded to construct the density matrix shown in Fig. 3d of ref. 1 (see also ref. 2). The degree of polarization is the difference between orthogonally polarized photon intensities, divided by the sum. The intensity of a polarized photon is determined from a density matrix by the sum probability of detection with another photon of equal or opposite polarization. Errors are determined by propagating the measurement errors used to construct the matrix. The polarization assessments from rectilinear, diagonal and circular bases are combined and errors compounded, to give an overall polarization with error.

zation splitting can be eradicated. Only these dots show behaviour suggestive of entanglement, including polarization correlation in all measurement bases, correlation insensitive to rotation of the detection basis, and appearance of off-diagonal elements in density matrices. Emission of entangled photon pairs by the dot, diluted by background light, is the only reasonable explanation for the observed behaviour.

Our approach for realizing dot sources of entangled photons<sup>1</sup> is reconfirmed by using devices with reduced background<sup>3</sup>, which increase the eigenvalue significantly to

$0.72 \pm 0.02$ , without background subtraction. The tangle and concurrence are increased to  $0.19 \pm 0.03$  and  $0.44 \pm 0.03$ , respectively. Further improvement should be feasible with even better isolation of light emitted by the dots, by resonant electrical or optical excitation. Combined with recent advances in single-photon emitters, such as telecommunications wavelength operation, and improved efficiency, the future of quantum-dot entangled photon sources for real applications is promising.

**R. Mark Stevenson\***, **Robert J. Young\***†, **Paola Atkinson**†, **Ken Cooper**†,

**David A. Ritchie**†, **Andrew J. Shields**\*

\*Toshiba Research Europe, 260 Cambridge Science Park, Cambridge CB4 0WE, UK  
e-mail: mark.stevenson@crl.toshiba.co.uk

†Cavendish Laboratory, University of Cambridge, Cambridge CB3 0HE, UK

1. Stevenson, R. M. *et al.* *Nature* **439**, 179–182 (2006).
2. Gilchrist, A., Resch, K. J. & White, A. G. *Nature* **445**, doi: 10.1038/nature05546 (2007).
3. Young, R. J. *et al.* *N. J. Physics* **8**, 29 (2006).
4. Edamatsu, K., Oohata, G., Shimizu, R. & Itoh, T. *Nature* **431**, 167–170 (2004).

doi:10.1038/nature05547



# Spatial coupling of nitrogen inputs and losses in the ocean

Curtis Deutsch<sup>1</sup>, Jorge L. Sarmiento<sup>2</sup>, Daniel M. Sigman<sup>3</sup>, Nicolas Gruber<sup>4,†</sup> & John P. Dunne<sup>5</sup>

**Nitrogen fixation is crucial for maintaining biological productivity in the oceans, because it replaces the biologically available nitrogen that is lost through denitrification. But, owing to its temporal and spatial variability, the global distribution of marine nitrogen fixation is difficult to determine from direct shipboard measurements. This uncertainty limits our understanding of the factors that influence nitrogen fixation, which may include iron, nitrogen-to-phosphorus ratios, and physical conditions such as temperature. Here we determine nitrogen fixation rates in the world's oceans through their impact on nitrate and phosphate concentrations in surface waters, using an ocean circulation model. Our results indicate that nitrogen fixation rates are highest in the Pacific Ocean, where water column denitrification rates are high but the rate of atmospheric iron deposition is low. We conclude that oceanic nitrogen fixation is closely tied to the generation of nitrogen-deficient waters in denitrification zones, supporting the view that nitrogen fixation stabilizes the oceanic inventory of fixed nitrogen over time.**

The major algal nutrients  $\text{NO}_3^-$  and  $\text{PO}_4^{3-}$  are in short supply throughout the well-lit surface waters of the low-latitude ocean. Although  $\text{NO}_3^-$  is usually depleted before  $\text{PO}_4^{3-}$ , the N limitation of productivity may be overcome by organisms capable of converting abundant dissolved  $\text{N}_2$  into 'fixed' N available to the wider ecosystem<sup>1</sup>. In oligotrophic regions of the warm subtropical ocean,  $\text{N}_2$  fixation has been estimated to account for ~50% of the organic carbon exported from the photic zone<sup>2</sup>. However, a high iron requirement of the enzyme system nitrogenase is believed to prevent  $\text{N}_2$ -fixing organisms from alleviating widespread N limitation<sup>3</sup>. The iron supply from atmospheric dust deposited at the sea surface has been hypothesized to favour  $\text{N}_2$  fixation in the more continentally influenced Atlantic Ocean<sup>4</sup>, and during relatively dusty ice ages, potentially increasing the N reservoir and productivity of the ice age ocean and thus lowering  $\text{CO}_2$  in the ice age atmosphere (ref. 5).

The environmental controls on  $\text{N}_2$  fixation and thus its probable response to past and future climate change would be much clearer if we knew the geographic distribution of this process in the modern ocean. Moreover, the persistent question of whether the sources and sinks of fixed N are in balance<sup>6</sup> awaits a robust estimate for the globally integrated rate of oceanic  $\text{N}_2$  fixation. Here we determine both the distribution and global rate of  $\text{N}_2$  fixation by taking advantage of one of its signatures in surface waters: a net biological uptake of  $\text{PO}_4^{3-}$  that occurs in the absence of a stoichiometric uptake of  $\text{NO}_3^-$ . By interpreting the observed nutrient distributions in the context of an ocean circulation model we find a global rate of  $\text{N}_2$  fixation of  $\sim 140 \times 10^{12}$  g of N per year occurring disproportionately in the Pacific Ocean.

## Nutrient fluxes and $\text{N}_2$ fixation

The biological uptake and remineralization of  $\text{NO}_3^-$  and  $\text{PO}_4^{3-}$  in the ocean are traditionally assumed to occur with a nearly constant ratio  $r_n$  of 16:1 (ref. 1). This assumption, while valid in the global mean, is clearly violated in the presence of  $\text{N}_2$  fixation, which requires  $\text{PO}_4^{3-}$  uptake but no concomitant uptake of fixed N. In regions where  $\text{N}_2$  fixation occurs, the consumption of  $\text{PO}_4^{3-}$

unaccompanied by fixed N will cause overall surface nutrient draw-down to have an N:P ratio below  $r_n$ , leaving the residual surface nutrient pool increasingly depleted in  $\text{PO}_4^{3-}$  relative to  $\text{NO}_3^-$ . A convenient measure of the depletion of  $\text{PO}_4^{3-}$  relative to the biological  $\text{NO}_3^-$  requirement is given by  $P^* = \text{PO}_4^{3-} - \text{NO}_3^-/r_n$ , which represents the excess of P relative to the standard N quota (see ref. 7 for a related term,  $N^*$ , describing N excess). While nutrient uptake by non- $\text{N}_2$ -fixing organisms will on average consume  $\text{NO}_3^-$  and  $\text{PO}_4^{3-}$  in a proportion that conserves  $P^*$ ,  $\text{N}_2$  fixation will extract  $\text{PO}_4^{3-}$  alone, driving a water parcel towards lower  $P^*$ .

According to this simple conceptual model,  $\text{N}_2$  fixation will be revealed as a reduction in  $P^*$  along the transport path of a surface water mass, and its rate can be estimated by combining the observed distributions of nutrients with information about the rate of ocean circulation and mixing. In the steady state, the excess uptake of  $\text{PO}_4^{3-}$  is equal to the transport convergence of  $P^*$ , written as  $-\nabla \cdot \Phi(P^*)$ , where  $\Phi$  represents nutrient fluxes resulting from all physical transport processes. When the circulation and mixing of the upper ocean results in an excess  $\text{PO}_4^{3-}$  supply,  $-\nabla \cdot \Phi(P^*) > 0$ , a biological uptake of  $\text{PO}_4^{3-}$  beyond the standard N quota, will be required to bring  $\text{NO}_3^-$  and  $\text{PO}_4^{3-}$  back towards their observed concentrations (Fig. 1). To the degree that non- $\text{N}_2$ -fixing phytoplankton produce most of the organic matter that actually sinks out of the surface ocean, the rate of  $\text{N}_2$  fixation associated with a given excess P supply is well approximated by  $J_{\text{fix}}(\text{N}) \approx -r_n \nabla \cdot \Phi(P^*)$ . However, an exact expression requires two modifications, one that accounts for any direct export of organic matter from  $\text{N}_2$ -fixing organisms with a high-biomass N:P ratio, and another that incorporates the effect of dissolved organic matter (DOM) cycling on the inferred supply of N and P.

The best-studied and perhaps dominant  $\text{N}_2$  fixer in the ocean, *Trichodesmium* spp., has an N:P ratio  $r_f$  that is several-fold higher than that of non- $\text{N}_2$ -fixing algae (ref. 8). A small fraction of *Trichodesmium* biomass may be directly exported as sinking particles rather than being recycled in surface waters, increasing the amount of newly fixed N that should be attributed to P consumption by

<sup>1</sup>Program on Climate Change, School of Oceanography, University of Washington, Seattle, Washington 98195, USA. <sup>2</sup>Atmospheric and Oceanic Sciences Program, Princeton University, Princeton, New Jersey 08544, USA. <sup>3</sup>Department of Geosciences, Guyot Hall, Princeton University, Princeton, New Jersey 08544, USA. <sup>4</sup>IGPP and Department of Atmospheric and Oceanic Sciences, University of California at Los Angeles, Los Angeles, California 90095, USA. <sup>5</sup>NOAA/Geophysical Fluid Dynamics Laboratory, PO Box 308, Forrestal Campus B Site, Princeton, New Jersey 08542, USA. <sup>†</sup>Present address: Institute of Biogeochemistry and Pollutant Dynamics, ETH Zurich, Zurich, Switzerland.

*Trichodesmium*. The relationship between the physical supply of excess  $\text{PO}_4^{3-}$  and the inferred rate of  $\text{N}_2$  fixation can be generalized (see Methods) as  $J_{\text{fix}}(\text{N}) = -\lambda \nabla \cdot \Phi(\text{P}^*)$ , where  $\lambda$  is a non-dimensional parameter (Supplementary Fig. 1),  $\lambda = \frac{1 - \gamma_e(1 - r_n/r_f)}{1 - \gamma_e(1 - r_n/r_f)}$ , and where  $\gamma_e$  is the fraction of organic matter production that is exported as sinking particles. If all organic matter production is directly exported in sinking particles ( $\gamma_e = 1$ ), any excess physical  $\text{PO}_4^{3-}$  supply must be consumed and exported by  $\text{N}_2$  fixers so that  $\lambda$  is equal to their biomass N:P ratio  $r_f$  (Fig. 1a). In contrast, if nutrient recycling is very efficient ( $\gamma_e \ll 1$ , Fig. 1b), then newly fixed N added to surface waters makes the inferred excess  $\text{PO}_4^{3-}$  uptake less than the excess physical  $\text{PO}_4^{3-}$  supply, and  $\lambda$  approaches  $r_n$ , as anticipated above. Because of the high degree of nutrient recycling in the low- and mid-latitude ocean,  $\lambda$  is only slightly (0–20%) greater than  $r_n$  (Supplementary Fig. 2), and the estimate of  $\text{N}_2$  fixation from surface nutrients is only weakly sensitive to  $r_f$ . However,  $\text{N}_2$  fixation by diazotrophs with both a high-biomass N:P and an unusually high export efficiency, should they be found, would be underestimated by our approach. Given that the standard N:P ratio of 16:1 (the Redfield ratio) actually derives from measurements of total marine particulates<sup>2</sup> and dissolved nutrients<sup>9</sup>, an N:P uptake ratio for non- $\text{N}_2$ -fixing organisms must be somewhat lower than this to offset the high  $r_f$  from the small fraction of  $\text{N}_2$ -fixing biomass. We find that an  $r_n$  of 15:1 yields a global bulk biomass with a Redfield ratio of 16:1.

The disappearance rate of excess P must also be modified to account for the production and subsequent degradation of dissolved organic N (DON) and P (DOP) whose substantial surface reservoirs constitute a second, slower pathway of nutrient recycling over

months to years. The impact of DON and DOP on inferred  $\text{N}_2$  fixation rates is accommodated (see Methods) by considering the conservation of total fixed N ( $\text{N}_t = \text{NO}_3^- + \text{DON}$ ; excluding  $\text{NH}_4^+$  and  $\text{NO}_2^-$ , which are scarce) and total P ( $\text{P}_t = \text{PO}_4^{3-} + \text{DOP}$ ), whereby  $\text{P}^*$  is simply replaced by total  $\text{P}^*$  ( $\text{P}_t^*$ ), the sum of  $\text{P}^*$  and an analogous  $\text{DOP}^*$  ( $\text{P}_t^* = \text{P}^* + \text{DOP}^* = \text{P}_t - \text{N}_t/r_n$ ). Thus, attributing excess  $\text{PO}_4^{3-}$  uptake to  $\text{N}_2$ -fixing organisms implies that:

$$J_{\text{fix}}(\text{N}) = -\lambda \nabla \cdot \Phi(\text{P}_t^*)$$

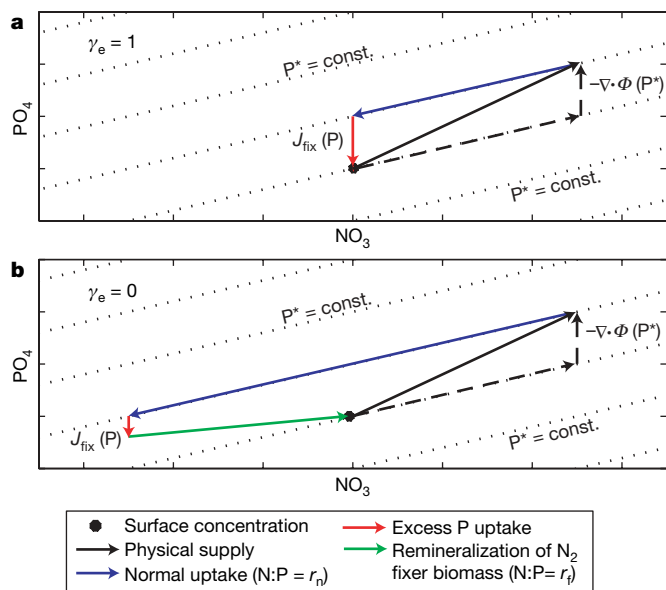
wherever the supply of total P exceeds that of total N times  $r_n$  (that is, where  $-\nabla \cdot \Phi(\text{P}_t^*) > 0$ ).

The distributions of  $\text{NO}_3^-$  and  $\text{PO}_4^{3-}$ , and thus of  $\text{P}^*$ , are relatively well known<sup>10</sup>. Except in the high-latitude ocean regions, in particular, the Southern Ocean, similar  $\text{P}^*$  patterns are observed in multiple data sets and therefore represent robust observations (Supplementary Fig. 3). In contrast, the spatial coverage of DON and DOP measurements is insufficient to estimate the transport divergence of  $\text{DOP}^*$  from those data alone. Instead, we simulate cycles of DON and DOP production and decay in an Ocean General Circulation Model (OGCM) that are driven by observed distributions of  $\text{NO}_3^-$  and  $\text{PO}_4^{3-}$  and independently tuned to match the patterns of DON and DOP where they have been most thoroughly studied (see Methods and Supplementary Table 1). The most critical observed pattern is of a tropical excess of DOP ( $\text{DOP}^* > 0$ ) and its decline towards a subtropical excess of DON ( $\text{DOP}^* < 0$ ) (ref. 11; Supplementary Fig. 4). This requires that, in the model, the fraction of  $\text{PO}_4^{3-}$  uptake converted to DOP is greater than the fractional production of DON, and that DOP is degraded more rapidly than DON<sup>12</sup>.

### Global distribution of $\text{N}_2$ fixation

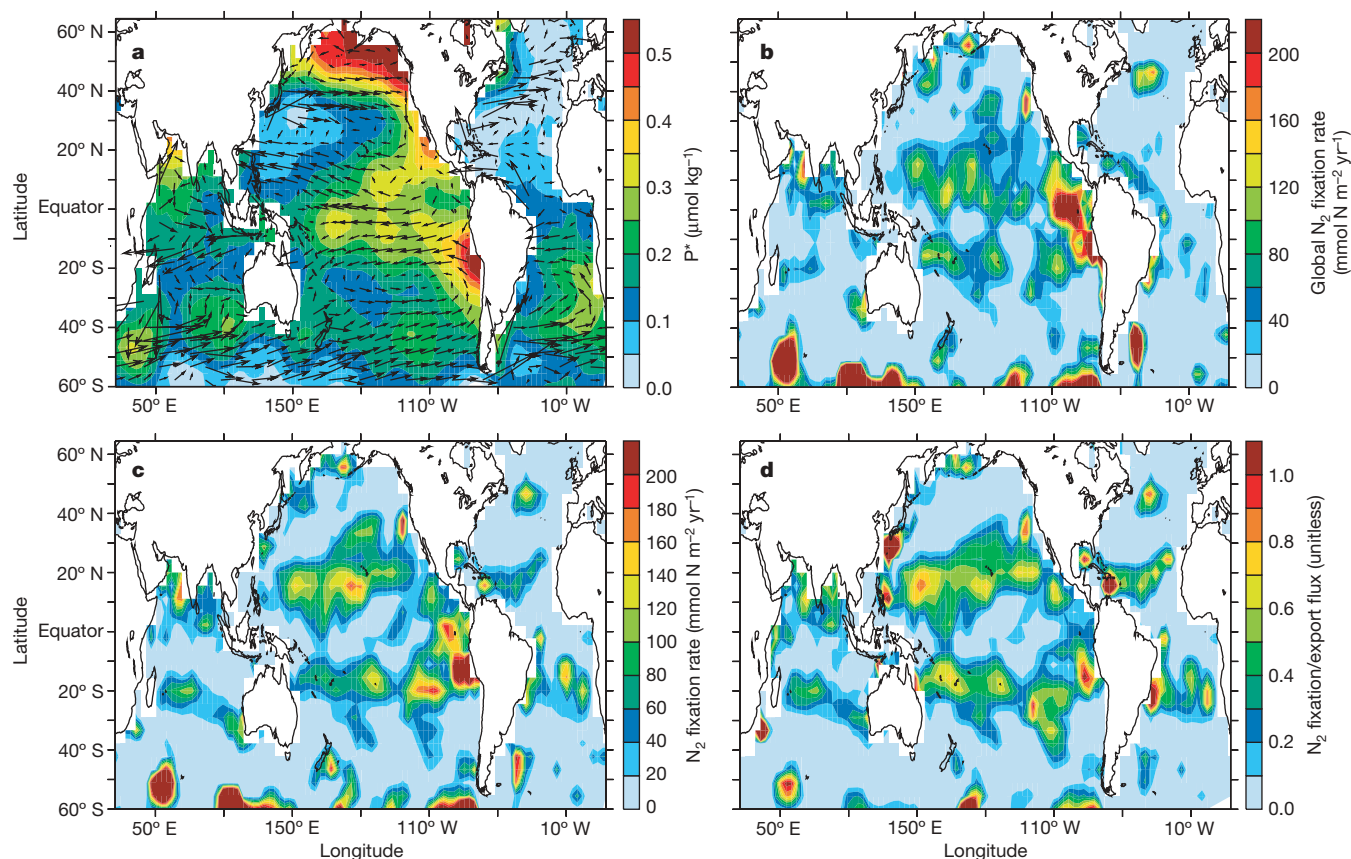
Having accounted for the surface ocean cycling of inorganic and organic nutrients,  $\text{N}_2$  fixation can be diagnosed from their physical fluxes computed by assimilating surface  $\text{NO}_3^-$  and  $\text{PO}_4^{3-}$  data in the OGCM (see Methods for more details). The climatological distribution of  $\text{P}^*$  (Fig. 2a) includes high excess  $\text{PO}_4^{3-}$  concentrations in the Eastern Tropical Pacific, the Subarctic North Pacific, the Arabian Sea, and off the western coast of Africa. In all of these regions,  $\text{NO}_3^-$  is consumed through denitrification (defined here as the sum of all processes that convert fixed N to  $\text{N}_2$ ) in subsurface waters, rendering them enriched in  $\text{PO}_4^{3-}$  relative to  $\text{NO}_3^-$ . Excess  $\text{PO}_4^{3-}$  declines towards zero in the subtropical gyre of the North Pacific, throughout the North Atlantic and, to a lesser degree, in the southern subtropical gyres of all three ocean basins. This  $\text{P}^*$  distribution, in the context of the OGCM circulation that transports high- $\text{P}^*$  waters into the lower- $\text{P}^*$  subtropical gyres, leads to diagnosed rates of  $\text{N}_2$  fixation that are focused in the low-latitude ocean (Fig. 2b, c). In the subtropical gyres, our results indicate that ~50% of the particulate N that is exported from the surface layer is newly fixed (as opposed to originating from nitrate supplied by the circulation, see Fig. 2d)<sup>3</sup>, making  $\text{N}_2$  fixation the dominant N supply that sustains biological carbon export in these systems. The latitudinal pattern of  $\text{N}_2$  fixation conforms broadly to the observed biogeography of *Trichodesmium* spp.<sup>13</sup>, which thrives in warm, low-nutrient waters. One inference from this result is that either *Trichodesmium* contributes most to global marine N inputs, or that environmental controls on the abundance of *Trichodesmium* are also operating on other  $\text{N}_2$ -fixing organisms<sup>14</sup>.

The cycling of DOM has little net effect on integrated  $\text{N}_2$  fixation rates (Table 1) mostly influencing its regional distribution within each basin (Fig. 2c compared to Fig. 2b). The relatively rapid production and degradation of DOP reduces the inferred excess of gross  $\text{PO}_4^{3-}$  uptake in productive regions while enhancing it in the 'downstream' oligotrophic regions, effectively shifting inferred  $\text{N}_2$  fixation away from the tropical upwelling zones and into the subtropical gyres (Fig. 2c) compared to estimates that either neglect DOM (Fig. 2b) or



**Figure 1 | Schematic depiction of method to determine  $\text{N}_2$  fixation from  $\text{NO}_3^-$  and  $\text{PO}_4^{3-}$  uptake.** Nutrient supply due to ocean circulation (black arrow) with an N:P ratio below  $r_n$  (dotted lines), indicates an excess supply of  $\text{PO}_4^{3-}$  ( $-\nabla \cdot \Phi(\text{P}^*) > 0$ , vertical dashed arrow) compared to normal biological N:P requirements (dashed arrow with constant  $\text{P}^*$ ). The total nutrient uptake required to balance the physical nutrient supply can be interpreted as a non- $\text{N}_2$ -fixing component that occurs with an N:P ratio of  $r_n$  (blue arrow), plus an excess of  $\text{PO}_4^{3-}$  uptake (red arrow) due to  $\text{N}_2$ -fixing organisms. When all nutrient uptake is rapidly exported (a), each mole of excess P uptake by  $\text{N}_2$  fixers carries  $r_f$  moles of newly fixed N to depth without altering the surface N supply. However, when nutrients are recycled in surface waters (b), the newly fixed N added to surface waters (green arrow) supplements the physical N supply, effectively reducing the excess  $\text{PO}_4^{3-}$  supply. The effect of nutrient recycling on  $\text{N}_2$  fixation rates inferred from a given nutrient supply depends on  $r_f$ , assumed here to be 50:1, and the fraction of total nutrient uptake exported in sinking particles ( $\gamma_e$ ), assumed to be 1 in a and 0 in b. The combined effect of these parameters is represented by  $\lambda$ .





**Figure 2 | Annual mean distribution of  $P^*$ , ocean currents, and the  $N_2$  fixation rates determined from them at 0–120 m depth. a, The  $P^*$  distribution ( $P^* = PO_4^{3-} - NO_3^-/r_n$ ) is based on climatological data from the World Ocean Atlas<sup>10</sup>, and the surface velocity is computed from the MOM3 ocean general circulation model<sup>30</sup>. b, Global  $N_2$  fixation rates diagnosed from the convergence of excess inorganic  $PO_4^{3-}$ ,  $-\lambda \nabla \cdot \Phi(P^*)$ ,**

which requires an excess uptake of  $PO_4^{3-}$  relative to the biological N requirement. c, Rates of  $N_2$  fixation accounting for both inorganic and organic nutrient pools, equal to  $-\lambda \nabla \cdot \Phi(P_i^*)$  where this term is positive (that is, where excess  $P_i$  converges). d,  $N_2$  fixation rates (from c) as a fraction of the export flux of organic matter.

assume proportional production and decay of DON and DOP (Supplementary Fig. 5). On the basis of the complete calculation including the DOM cycle (Fig. 2c),  $N_2$  fixation appears to be proximal to but does not overlap with regions of non-zero nutrients. However, given data coverage and the likelihood of unresolved seasonality, we avoid using its detailed distribution to address whether  $N_2$  fixation can occur in nutrient-replete waters.

The occurrence of maximum inferred  $N_2$  fixation rates in tropical and subtropical latitudes is highly consistent with previous views<sup>13</sup>, but its inter-basin distribution is not. Whereas there is a widely held

expectation of higher rates in the Atlantic<sup>4</sup>, our estimates imply areal rates of  $N_2$  fixation in the Pacific Ocean more than twice as high as those in the Atlantic, with intermediate values in the Indian Ocean (Table 1). The Pacific hosts over two-thirds of the total  $\sim 140$  Tg ( $1 \text{ Tg} = 10^{12} \text{ g}$ ) of new N required each year to account for the excess uptake of  $PO_4^{3-}$ , followed by 16% of the total in the Indian Ocean. Denitrification in the eastern tropical Pacific<sup>15</sup>, subarctic North Pacific<sup>16</sup> and Arabian Sea<sup>17</sup> generates nutrient-rich and  $NO_3^-$ -deficient waters that subsequently undergo a disproportionate loss of  $PO_4^{3-}$  as they are upwelled and transported into the adjacent subtropical gyres. In contrast, the relatively weak  $P^*$  gradients found across much of the subtropical North Atlantic require nutrient uptake with N:P ratios that are close to or even greater than  $r_n$ . Our basin-wide  $N_2$  fixation rate for the Pacific ( $48 \text{ mmol N m}^{-2} \text{ yr}^{-1}$ , averaged from  $40^\circ \text{ S}$  to  $65^\circ \text{ N}$ ) is similar to previous estimates ( $\sim 50 \text{ mmol N m}^{-2} \text{ yr}^{-1}$ ; refs 1, 18). In the Atlantic, our estimate of  $23 \text{ mmol N m}^{-2} \text{ yr}^{-1}$  is at the low end of the range of previous studies based on thermocline nutrient stoichiometry (30 to  $72 \text{ mmol N m}^{-2} \text{ yr}^{-1}$ ; refs 7, 19), and is also less than a recent estimate ( $87 \text{ mmol N m}^{-2} \text{ yr}^{-1}$ ) from *in situ* measurements in the western tropical North Atlantic<sup>20</sup>, although these studies examined a more selective area, where average rates can be expected to be higher.

We have investigated the sensitivity of our results to uncertainties in ocean circulation, the distribution of  $P^*$ , and the N:P ratio of non- $N_2$  fixing plankton,  $r_n$ . We find that globally integrated  $N_2$  fixation rates among the sensitivity experiments vary by less than 25% (Table 1). While the details of the inferred distribution of  $N_2$  fixation may vary, the partitioning between ocean basins is robust, with the Atlantic consistently contributing 15% or less to the global input

**Table 1 | Global and basin-scale  $N_2$  fixation rates diagnosed from surface nutrients**

Model description	$N_2$ fixation rate ( $10^{12} \text{ g N yr}^{-1}$ )		
	Pacific ( $130 \times 10^{12} \text{ m}^2$ )	Atlantic ( $63 \times 10^{12} \text{ m}^2$ )	Global ( $240 \times 10^{12} \text{ m}^2$ )
Standard Model	95	20	137
No DOM	95	15	130
Alternative $P^*$ (WOCE)	93	21	143
Alternative circulation (winds)	108	17	151
Alternative circulation (mixing)	98	13	140
Assume $r_n = 16$	107	25	158

Rates are integrated over the top 120 m from  $40^\circ \text{ S}$  to  $65^\circ \text{ N}$  for several models: the Standard Model, a model without DOM, and four sensitivity experiments (see Supplementary Information), in which the effect of uncertainty in physical nutrient transport is estimated using an alternative  $P^*$  distribution (based on quality-controlled nutrient data from the World Ocean Circulation Experiment, WOCE) and two alternative circulations (one that is forced by an alternate pattern of wind stress, and one that additionally imposes higher rates of mixing). These calculations, together with one in which the N:P uptake ratio for non- $N_2$ -fixing plankton ( $r_n$ ) is assumed to be 16:1, reveal a pattern and global rate of  $N_2$  fixation that is robust with respect to the key methodological uncertainties.

(Supplementary Figs 7, 8). The greatest dependence of the global  $N_2$  fixation rate is on  $r_n$ . For a given ratio of total nutrient drawdown, a higher N:P ratio among non- $N_2$ -fixing organisms requires a greater fraction of P uptake to be attributed to diazotrophs (Table 1); however, the large-scale distribution of inferred  $N_2$  fixation is not changed.

### Internal regulation of oceanic fixed N

The pattern of  $N_2$  fixation diagnosed from surface nutrients (Fig. 2c) has several important implications for the understanding of environmental controls on marine  $N_2$  fixation. The inter-basin differences in diagnosed  $N_2$  fixation rates are opposite to those for iron supplied through atmospheric dust deposition<sup>21</sup>, which is higher in the Atlantic than in the Pacific. This suggests that the supply of iron may not be a primary limiting factor for marine  $N_2$  fixation, or that the supply of iron from the subsurface compensates for low atmospheric inputs in the Pacific. Instead, regions of  $N_2$  fixation diagnosed from surface nutrient transport are closely connected to zones of low  $O_2$  and active denitrification in the water column<sup>17,22</sup>. This suggests that  $NO_3^-$  deficits generated in the suboxic zones of the Pacific and Indian Oceans and subsequently transported to surface waters provide an important stimulus for  $N_2$  fixation, perhaps by increasing the ability of  $N_2$  fixers to compete with other phytoplankton.

The relatively tight coupling between denitrification and  $N_2$  fixation inferred from this analysis in turn offers support for a previously hypothesized negative feedback in the N budget, whereby any externally driven increase or decrease in the oceanic N reservoir will be counteracted by a decrease or increase in the rate of  $N_2$  fixation, respectively<sup>1,23</sup>. For example, our results support the hypothesis that  $N_2$  fixation responded to the reconstructed surge in water column denitrification at the end of the last ice age by increasing in step<sup>24</sup>. Moreover, given the close geographic proximity of denitrification and  $N_2$  fixation noted here, this response could have occurred on a timescale of years to decades, which is nearly instantaneous on the timescale of glacial–interglacial transitions. Finally, this compensation would occur within individual ocean basins, not requiring that, for instance, changes in Pacific denitrification were countered by comparably large changes in Atlantic  $N_2$  fixation.

The proximity of denitrification and  $N_2$  fixation centres suggests that the stabilizing N-to-P feedback would have been operating over recent decades while the ocean N budget has been studied. If so, this argues against the interpretation that the recognized difference between the estimated inputs and outputs of oceanic fixed N could be explained as high-frequency transient imbalances within a longer-term steady state<sup>25</sup>. Rather, our results argue against any such imbalance. At the same time, our globally integrated estimate for oceanic  $N_2$  fixation, like previous estimates, is far below many current estimates for total denitrification (for example,  $\sim 350$  Tg of N per year<sup>26</sup>). Taking our results at face value, either the estimates for fixed N loss must be too large, or we must be missing a major additional input of N to the ocean.

### METHODS

**Nutrient mass balance.** The rate of  $N_2$  fixation diagnosed from surface nutrients can be generalized by considering the conservation of inorganic nutrients and DOM in surface waters:

$$\frac{\partial P}{\partial t} + \nabla \cdot \Phi(P) = J_{DOM}(P) - J_{up}(P) + J_{regen}(P) \quad (1)$$

$$\frac{\partial N}{\partial t} + \nabla \cdot \Phi(N) = J_{DOM}(N) - J_{up}(N) + J_{regen}(N) \quad (2)$$

$$\frac{\partial DOP}{\partial t} + \nabla \cdot \Phi(DOP) = -J_{DOM}(P) + \gamma_P J_{up}(P) \quad (3)$$

$$\frac{\partial DON}{\partial t} + \nabla \cdot \Phi(DON) = -J_{DOM}(N) + \gamma_N [J_{fix}(N) + J_{up}(N)] \quad (4)$$

where P and N are the respective concentrations of inorganic  $PO_4^{3-}$  and inorganic fixed N, which is dominated by  $NO_3^-$  ( $N = NH_4^+ + NO_2^- +$

$NO_3^- \approx NO_3^-$ ). The  $\nabla \cdot \Phi$  terms represent the divergence of physical fluxes due to advection and diffusion, and the  $J$  terms represent biological sources and sinks of nutrients.  $J_{up}$  is total nutrient uptake including that associated with  $N_2$  fixation,  $J_{DOM}$  is the degradation of DON and DOP,  $J_{regen}$  is the instantaneous regeneration of surface nutrients, and  $J_{fix}(N)$  is the fixation of  $N_2$  into organic N. A fraction  $\gamma_N$  of total primary N production,  $J_{fix}(N) + J_{up}(N)$ , is converted to DON, while a distinct fraction of gross P uptake  $\gamma_P$  is converted to DOP. The fraction of organic matter production that sinks as particulate organic matter is denoted  $\gamma_e$ , so that the regenerated nutrient flux is what remains:

$$J_{regen}(P) = (1 - \gamma_e - \gamma_P) J_{up}(P) \quad (5)$$

$$J_{regen}(N) = (1 - \gamma_e - \gamma_N) (J_{up}(N) + J_{fix}(N)) \quad (6)$$

In contrast to P, surface N regeneration and DON production may introduce exogenous, newly fixed N from  $N_2$  fixation. Equations (1)–(5) can be combined to describe the conservation of total phosphorus ( $P_t = P + DOP$ ), and total fixed nitrogen ( $N_t = N + DON$ ):

$$\frac{\partial P_t}{\partial t} + \nabla \cdot \Phi(P_t) = -\gamma_e J_{up}(P) \quad (7)$$

$$\frac{\partial N_t}{\partial t} + \nabla \cdot \Phi(N_t) = -\gamma_e J_{up}(N) + (1 - \gamma_e) J_{fix}(N) \quad (8)$$

Equations (7) and (8) state that the time rate-of-change and transport divergence of  $P_t$  is balanced by the sinking flux (export) of P, and similarly for  $N_t$ , except that there is also a source of  $N_t$  from the fraction of newly fixed N that does not get exported.

Our fundamental assumption is that  $N_2$  fixation can be estimated from (see Fig. 1):

$$J_{fix}(N) = r_f (J_{up}(P) - \frac{J_{up}(N)}{r_n}) \quad (9)$$

where  $r_n$  is the N:P ratio of ‘normal’ nutrient uptake and  $r_f$  is the N:P ratio of  $N_2$  fixers. Therefore,  $N_2$  fixation can be diagnosed (in the steady state) using equations (7)–(9) from:

$$J_{fix}(N) = -\lambda \nabla \cdot \Phi(P_t^*) \text{ only if } -\nabla \cdot \Phi(P_t^*) > 0 \quad (10)$$

where  $-\nabla \cdot \Phi(P_t^*)$  is the convergence of  $P_t^*$  due to circulation, and  $\lambda = r_n [1 - \gamma_e (1 - r_n/r_f)]^{-1}$  can be thought of as the amount of N that must be fixed to compensate a given excess physical supply of  $PO_4^{3-}$ . Its dependence on  $r_f$  has two opposing effects. On the one hand, the amount of N fixed per unit of  $PO_4^{3-}$  uptake by  $N_2$  fixers increases with  $r_f$  so that, for a higher  $r_f$ , a given burden of excess  $PO_4^{3-}$  supplied by the circulation is associated with a higher  $N_2$  fixation rate. On the other hand, as  $r_f$  increases, the net excess of  $PO_4^{3-}$  supplied to the surface ocean decreases owing to higher *in situ* N inputs from  $N_2$  fixers. Thus, the amount of  $N_2$  fixation required to balance a given supply of excess  $PO_4^{3-}$  inferred from nutrient distributions and ocean flow fields depends on both the N:P ratio of  $N_2$ -fixing biomass (an increase in the N:P ratio will increase the amount of  $N_2$  fixed) and the degree of nutrient recycling (an increase in which will reduce the amount of  $N_2$  fixed, and more so for a higher  $r_f$ ). Because the expression inside the square brackets is close to 1, the value of  $\lambda$  tends to be close to that of  $r_n$  (Supplementary Fig. 2).

**Ocean circulation/biogeochemistry model.** We compute physical and biological nutrient fluxes within an ocean biogeochemical/general circulation model that integrates the conservation equations (1)–(4) to a steady state. Model  $NO_3^-$  and  $PO_4^{3-}$  concentrations are damped towards climatological values (Fig. 2a), providing an estimate of gross nutrient uptake rates. An excess uptake of P is attributed to  $N_2$ -fixing organisms with an N:P ratio of 50:1; see equation (9). We assume that regions with an excess of N supply relative to P (that is,  $-\nabla \cdot \Phi(P_t^*) < 0$ ; Supplementary Fig. 8) are balanced by high N:P uptake<sup>27</sup>. An alternative explanation would be that biological nutrient uptake occurs at the standard ratio but that an additional process, such as denitrification, removes N but not P. This could occur in regions where denitrifying sediments are found in the shallow water column, for example, in the subarctic North Pacific<sup>28</sup>.

The fraction of total productivity that is exported in sinking particles ( $\gamma_e$ ) is determined through an empirical model in which  $\gamma_e$  is an increasing function of productivity and a decreasing function of surface temperature (Supplementary Fig. 2)<sup>29</sup>. DON and DOP are produced in proportion to N and P uptake via the parameters  $\gamma_N$  and  $\gamma_P$ , and consumed in a first-order decay process (that is,  $J_{DOM}(N) = DON/\tau_{DON}$  and  $J_{DOM}(P) = DOP/\tau_{DOP}$ ; see the ‘Nutrient mass balance’ section of the Methods) that is tuned to match the available observations (see Supplementary Table 1). Tropical/subtropical DON and DOP maxima are most accurately reproduced when the fractional production of DON and DOP ( $\gamma_N$  and  $\gamma_P$ ) are 6% and 25%, respectively, and their degradation timescales



( $\tau_{\text{DON}}$  and  $\tau_{\text{DOP}}$ ) are 2 and 0.5 years, respectively. By tuning the parameters for DOM cycling and restoring N and P distributions towards observations, we ensure that the transport of N, P, DON and DOP is maximally consistent with observed values. Tracer transport is computed within a coarse-resolution ( $\sim 4^\circ$ ) general circulation model, the Modular Ocean Model (MOM3). We adopt here the model configuration P2A, in which both the upper-ocean thermal structure and global new production rates are consistent with observational constraints<sup>30</sup>.

Received 7 August; accepted 7 November 2006.

- Redfield, A. C., Ketchum, B. H. & Richards, F. A. in *The Sea* (ed. Hill, M. N.) Vol. 2, 26–77 (Interscience, New York, 1963).
- Karl, D. et al. The role of nitrogen fixation in biogeochemical cycling in the subtropical North Pacific Ocean. *Nature* **388**, 533–538 (1997).
- Falkowski, P. G. Evolution of the nitrogen cycle and its influence on the biological sequestration of  $\text{CO}_2$  in the ocean. *Nature* **387**, 272–275 (1997).
- Karl, D. et al. Dinitrogen fixation in the world's oceans. *Biogeochemistry* **57/58**, 47–98 (2002).
- Broecker, W. S. & Henderson, G. M. The sequence of events surrounding Termination II and their implications for the cause of glacial-interglacial  $\text{CO}_2$  changes. *Paleoceanography* **13**, 352–364 (1998).
- Codispoti, L. A. Biogeochemical cycles—Is the ocean losing nitrate? *Nature* **376**, 724 (1995).
- Gruber, N. & Sarmiento, J. L. Global patterns of marine nitrogen fixation and denitrification. *Glob. Biogeochem. Cycles* **11**, 235–266 (1997).
- Letelier, R. M. & Karl, D. M. *Trichodesmium* spp. physiology and nutrient fluxes in the North Pacific subtropical gyre. *Aquat. Microb. Ecol.* **15**, 265–276 (1998).
- Anderson, L. A. & Sarmiento, J. L. Redfield ratios of remineralization determined by nutrient data analysis. *Glob. Biogeochem. Cycles* **8**, 65–80 (1994).
- Conkright, M. E. et al. *World Ocean Atlas 2001: Objective Analyses, Data Statistics, and Figures, CD-ROM Documentation 1–17* (National Oceanographic Data Center, Silver Spring, 2002).
- Abell, J., Emerson, S. & Renaud, P. Distributions of TOP, TON and TOC in the North Pacific subtropical gyre: Implications for nutrient supply in the surface ocean and remineralization in the upper thermocline. *J. Mar. Res.* **58**, 203–222 (2000).
- Wu, J. F., Sunda, W., Boyle, E. A. & Karl, D. M. Phosphate depletion in the western North Atlantic Ocean. *Science* **289**, 759–762 (2000).
- Capone, D. G., Zehr, J. P., Paerl, H. W., Bergman, B. & Carpenter, E. J. *Trichodesmium*, a globally significant marine cyanobacterium. *Science* **276**, 1221–1229 (1997).
- Zehr, J. P. et al. Unicellular cyanobacteria fix  $\text{N}_2$  in the subtropical North Pacific Ocean. *Nature* **412**, 635–638 (2001).
- Codispoti, L. A. & Richards, F. A. An analysis of the horizontal regime of denitrification in the eastern tropical North Pacific. *Limnol. Oceanogr.* **21**, 379–388 (1976).
- Lehmann, M. F. et al. Origin of the deep Bering Sea nitrate deficit: Constraints from the nitrogen and oxygen isotopic composition of water column nitrate and benthic nitrate fluxes. *Glob. Biogeochem. Cycles* **19**, doi:10.1029/2005GB002508 (2005).
- Brandes, J. A., Devol, A. H., Yoshinari, T., Jayakumar, D. A. & Naqvi, S. W. A. Isotopic composition of nitrate in the central Arabian Sea and eastern tropical North Pacific: a tracer for mixing and nitrogen cycles. *Limnol. Oceanogr.* **43**, 1680–1689 (1998).
- Deutsch, C., Gruber, N., Key, R. M., Sarmiento, J. L. & Ganaschaud, A. Denitrification and  $\text{N}_2$  fixation in the Pacific Ocean. *Glob. Biogeochem. Cycles* **15**, 483–506 (2001).
- Hansell, D. A., Bates, N. R. & Olson, D. B. Excess nitrate and nitrogen fixation in the North Atlantic Ocean. *Mar. Chem.* **84**, 243–265 (2004).
- Capone, D. G. et al. Nitrogen fixation by *Trichodesmium* spp.: An important source of new nitrogen to the tropical and subtropical North Atlantic Ocean. *Glob. Biogeochem. Cycles* **19**, doi:10.1029/2004GB002331 (2005).
- Mahowald, N. et al. Dust sources and deposition during the last glacial maximum and current climate: A comparison of model results with paleodata from ice cores and marine sediments. *J. Geophys. Res. Atmos.* **104**, 15895–15916 (1999).
- Sigman, D. M. et al. Coupled nitrogen and oxygen isotope measurements of nitrate along the eastern North Pacific margin. *Glob. Biogeochem. Cycles* **19**, doi:10.1029/2005GB002458 (2005).
- Gruber, N. in *Carbon-Climate Interactions* (eds Follows, M. & Oguz, T.) 97–148 (John Wiley & Sons, New York, 2003).
- Deutsch, C., Sigman, D. M., Thunell, R., Meckler, A. N. & Haug, G. H. Isotopic Constraints on the Glacial/Interglacial Oceanic Nitrogen Budget. *Glob. Biogeochem. Cycles* **18**, doi:10.1029/2003GB002189 (2004).
- Codispoti, L. A. et al. The oceanic fixed nitrogen and nitrous oxide budgets: Moving targets as we enter the anthropocene? *Sci. Mar.* **65**, 85–101 (2001).
- Brandes, J. A. & Devol, A. H. A global marine fixed nitrogen isotopic budget: Implications for Holocene nitrogen cycling. *Glob. Biogeochem. Cycles* **16**, doi:10.1029/2001GB001856 (2002).
- Sarmiento, J. L. & Gruber, N. *Ocean Biogeochemical Dynamics* 118–119 (Princeton Univ. Press, Princeton, 2006).
- Tanaka, T. N deficiency in a well-oxygenated cold bottom water over the Bering Sea shelf: Influence of sedimentary denitrification. *Contin. Shelf Res.* **24**, 1271–1283 (2004).
- Dunne, J. P., Armstrong, R. A., Gnanadesikan, A. & Sarmiento, J. L. Empirical and mechanistic models for the particle export ratio. *Glob. Biogeochem. Cycles* **19**, doi:10.1029/2004GB002390 (2004).
- Gnanadesikan, A., Slater, R. D., Gruber, N. & Sarmiento, J. L. Oceanic vertical exchange and new production: a comparison between models and observations. *Deep-Sea Res. II* **49**, 363–401 (2002).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** C.D. was supported by a NASA Earth System Science Fellowship and the UW Program on Climate Change. J.L.S. and N.G. acknowledge support from the Office of Science (BER) and the US Department of Energy. J.L.S. also acknowledges support from the National Oceanic and Atmospheric Administration. D.M.S. acknowledges support from the US NSF.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to C.D. (cdeutsch@ocean.washington.edu).

## ARTICLES

# Genome-wide atlas of gene expression in the adult mouse brain

Ed S. Lein<sup>1\*</sup>, Michael J. Hawrylycz<sup>1\*</sup>, Nancy Ao<sup>2</sup>, Mikael Ayres<sup>1</sup>, Amy Bensinger<sup>1</sup>, Amy Bernard<sup>1</sup>, Andrew F. Boe<sup>1</sup>, Mark S. Boguski<sup>1†</sup>, Kevin S. Brockway<sup>1</sup>, Emi J. Byrnes<sup>1</sup>, Lin Chen<sup>1</sup>, Li Chen<sup>2</sup>, Tsuey-Ming Chen<sup>2</sup>, Mei Chi Chin<sup>1</sup>, Jimmy Chong<sup>1</sup>, Brian E. Crook<sup>1</sup>, Aneta Czaplinska<sup>2</sup>, Chinh N. Dang<sup>1</sup>, Suvro Datta<sup>1</sup>, Nick R. Dee<sup>1</sup>, Aimee L. Desaki<sup>1</sup>, Tsega Desta<sup>1</sup>, Ellen Diep<sup>1</sup>, Tim A. Dolbeare<sup>1</sup>, Matthew J. Donelan<sup>1</sup>, Hong-Wei Dong<sup>1</sup>, Jennifer G. Dougherty<sup>1</sup>, Ben J. Duncan<sup>1</sup>, Amanda J. Ebbert<sup>1</sup>, Gregor Eichele<sup>3</sup>, Lili K. Estin<sup>1</sup>, Casey Faber<sup>1</sup>, Benjamin A. Facer<sup>1</sup>, Rick Fields<sup>2</sup>, Shanna R. Fischer<sup>1</sup>, Tim P. Fliss<sup>1</sup>, Cliff Frensley<sup>1</sup>, Sabrina N. Gates<sup>1</sup>, Katie J. Glattfelder<sup>1</sup>, Kevin R. Halverson<sup>1</sup>, Matthew R. Hart<sup>1</sup>, John G. Hohmann<sup>1</sup>, Maureen P. Howell<sup>1</sup>, Darren P. Jeung<sup>1</sup>, Rebecca A. Johnson<sup>1</sup>, Patrick T. Karr<sup>1</sup>, Reena Kawal<sup>1</sup>, Jolene M. Kidney<sup>1</sup>, Rachel H. Knapik<sup>1</sup>, Chihchau L. Kuan<sup>1</sup>, James H. Lake<sup>1</sup>, Annabel R. Laramée<sup>1</sup>, Kirk D. Larsen<sup>1</sup>, Christopher Lau<sup>1</sup>, Tracy A. Lemon<sup>1</sup>, Agnes J. Liang<sup>2</sup>, Ying Liu<sup>2</sup>, Lon T. Luong<sup>1</sup>, Jesse Michaels<sup>1</sup>, Judith J. Morgan<sup>1</sup>, Rebecca J. Morgan<sup>1</sup>, Marty T. Mortrud<sup>1</sup>, Nerick F. Mosqueda<sup>1</sup>, Lydia L. Ng<sup>1</sup>, Randy Ng<sup>1</sup>, Geralyn J. Orta<sup>1</sup>, Caroline C. Overly<sup>1</sup>, Tu H. Pak<sup>1</sup>, Sheana E. Parry<sup>1</sup>, Sayan D. Pathak<sup>1</sup>, Owen C. Pearson<sup>1</sup>, Ralph B. Puchalski<sup>1</sup>, Zackery L. Riley<sup>1</sup>, Hannah R. Rockett<sup>1</sup>, Stephen A. Rowland<sup>1</sup>, Joshua J. Royall<sup>1</sup>, Marcos J. Ruiz<sup>2</sup>, Nadia R. Sarno<sup>1</sup>, Katherine Schaffnit<sup>1</sup>, Nadiya V. Shapovalova<sup>1</sup>, Taz Sivasay<sup>1</sup>, Clifford R. Slaughterbeck<sup>1</sup>, Simon C. Smith<sup>1</sup>, Kimberly A. Smith<sup>1</sup>, Bryan I. Smith<sup>1</sup>, Andy J. Sodt<sup>1</sup>, Nick N. Stewart<sup>1</sup>, Kenda-Ruth Stumpf<sup>1</sup>, Susan M. Sunkin<sup>1</sup>, Madhavi Sutram<sup>1</sup>, Angelene Tam<sup>2</sup>, Carey D. Teemer<sup>1</sup>, Christina Thaller<sup>2</sup>, Carol L. Thompson<sup>1</sup>, Lee R. Varnam<sup>1</sup>, Axel Visel<sup>3†</sup>, Ray M. Whitlock<sup>1</sup>, Paul E. Wohnoutka<sup>1</sup>, Crissa K. Wolkey<sup>1</sup>, Victoria Y. Wong<sup>1</sup>, Matthew Wood<sup>2</sup>, Murat B. Yaylaoglu<sup>2</sup>, Rob C. Young<sup>1</sup>, Brian L. Youngstrom<sup>1</sup>, Xu Feng Yuan<sup>1</sup>, Bin Zhang<sup>2</sup>, Theresa A. Zwingman<sup>1</sup> & Allan R. Jones<sup>1</sup>

**Molecular approaches to understanding the functional circuitry of the nervous system promise new insights into the relationship between genes, brain and behaviour. The cellular diversity of the brain necessitates a cellular resolution approach towards understanding the functional genomics of the nervous system. We describe here an anatomically comprehensive digital atlas containing the expression patterns of ~20,000 genes in the adult mouse brain. Data were generated using automated high-throughput procedures for *in situ* hybridization and data acquisition, and are publicly accessible online. Newly developed image-based informatics tools allow global genome-scale structural analysis and cross-correlation, as well as identification of regionally enriched genes. Unbiased fine-resolution analysis has identified highly specific cellular markers as well as extensive evidence of cellular heterogeneity not evident in classical neuroanatomical atlases. This highly standardized atlas provides an open, primary data resource for a wide variety of further studies concerning brain organization and function.**

The mammalian central nervous system (CNS) contains an enormous variety of cell types, each with unique morphology, connectivity, physiology and function. A characterization of the full complement of neural cell types is essential to understand functional circuit properties and their relation to higher cognitive functions and behaviours. The phenotypic properties of different neuronal and non-neuronal cells are largely the product of unique combinations of expressed gene products; therefore, gene expression profiles provide an informative modality to define cellular diversity in the brain. However, histological data are typically generated for one gene at a time, and data are neither systematically produced and analysed nor consolidated in an easily accessible format. Consequently, a limited set of

established cellular markers dominates the current literature, and expression patterns of many genes remain uncharacterized.

The recent availability of genomic sequence information and high-throughput advances in experimental techniques has allowed an expansion from single-gene to genome-wide transcriptional analyses. These approaches have been used to investigate the relationship between brain structure and function in normal adult brain, development and disease, most notably using DNA microarrays and serial analysis of gene expression (SAGE)<sup>1–4</sup>. However, these techniques have typically been applied to large brain regions, yielding data that are difficult to interpret because they do not resolve cellular diversity within those structures (although see refs 5 and 6). This diversity is

<sup>1</sup>Allen Institute for Brain Science, Seattle, Washington 98103, USA. <sup>2</sup>Verna and Marrs McLean Department of Biochemistry and Molecular Biology, Baylor College of Medicine, Houston, Texas 77030, USA. <sup>3</sup>Department of Genes and Behavior, Max Planck Institute of Biophysical Chemistry, 37077 Goettingen, Germany. <sup>†</sup>Present addresses: Division of Genome and Proteome Sciences, Novartis Institutes for Biomedical Research, 250 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA (M.S.B.); Lawrence Berkeley National Laboratory, Genomics Division, MS84-171, 1 Cyclotron Road, Berkeley, California 94720, USA (A.V.).

\*These authors contributed equally to this work.



exemplified by single neuron studies comparing gene expression with anatomical and physiological properties, demonstrating complex relationships between gene expression and neuronal subtypes, frequently with very different expression profiles in adjacent neurons<sup>7–9</sup>. The fine (cellular but not single cell as above<sup>7–9</sup>) resolution of scalable techniques such as *in situ* hybridization (ISH) is needed to begin to make meaningful global correlations between gene expression, gene regulation, CNS function and cellular phenotype. An increasing number of large-scale projects are now determining messenger RNA and protein localization in adult and developing brain<sup>10–14</sup>. The marriage of the unbiased genomic approach with semi-quantitative histological methods promises to open many new research avenues, particularly in a core model for mammalian brain development and behavioural genetics, the mouse<sup>15</sup>.

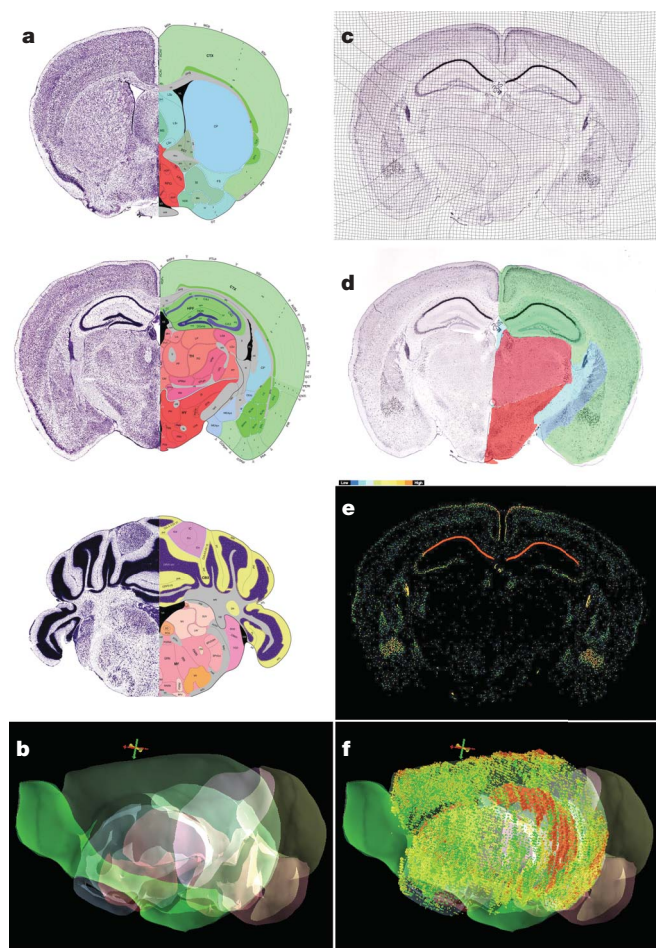
The Allen Brain Atlas project has taken a global approach to understanding the genetic structural and cellular architecture of the mouse brain by generating a genome-scale collection of cellular resolution gene expression profiles using ISH. Highly methodical data production methods and comprehensive anatomical coverage via dense, uniformly spaced sampling facilitate data consistency and comparability across >20,000 genes. The use of an inbred mouse strain with minimal animal-to-animal variance allows one to treat the brain

essentially as a complex but highly reproducible three-dimensional tissue array. Image-based informatics methods for signal detection and three-dimensional registration of ISH data to a *de novo* age-matched annotated reference atlas have been developed to allow automated signal quantification across the anatomical structures in the reference atlas, or its associated grid-based coordinate system. These methods enable global analysis and mining for detailed expression patterns in the brain. The entire Allen Brain Atlas data set and associated informatics tools are available through an unrestricted web-based viewing application (<http://www.brain-map.org>).

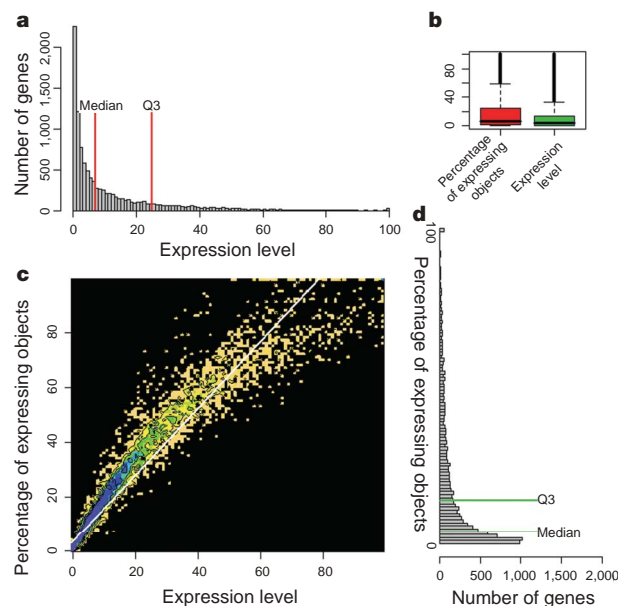
### Global analysis of gene expression

**Quantification and mapping strategies.** To create a searchable, anatomy-centred database of gene expression, ISH data for each gene is aligned in the same three-dimensional coordinate space through registration with a reference atlas created for the project (Supplementary Methods 3). Three-dimensional volumetric representations are created from the reference atlas and the series of ISH images for each gene (Fig. 1). The atlas volume is registered in three dimensions to the ISH volume, thereby alleviating animal-to-animal differences in cutting plane. This process applies a standardized three-dimensional grid coordinate frame to the ISH data for each gene (Fig. 1c), with mapping to corresponding anatomical structures provided by the annotated reference atlas (Fig. 1d). Although there are caveats associated with handling image artefacts and consistent registration across different brains, the result of this registration system is a genome-scale ISH data set searchable either by anatomical structure or using a finer (100  $\mu$ m to 300  $\mu$ m) resolution geometric grid.

Informatics tools have also been developed to identify pixels corresponding to true ISH signal and to identify (segment) expressing objects (contiguous clusters of pixels corresponding to cells) on the two-dimensional images for each gene. This process allows relative quantification based on either the fraction of expressing cells (relative density) or on expression level (approximating total transcript count),



**Figure 1 | Global analysis strategy.** **a**, Panels from the Allen Reference Atlas, annotated and colour-coded using an anatomical nomenclature modified from ref. 55. **b**, These images are rendered into a three-dimensional annotated reference volume. ISH images are aligned in three dimensions with the reference volume (*Man1a* shown in **c–f**), allowing overlay of a fine grid framework (**c**) or a plane-matched slice through the reference volume to delineate structural boundaries (**d**) onto the two-dimensional ISH images. **e**, **f**, Image segmentation and false-colouring by object pixel intensity (**e**) allows quantification, assignment of signal to anatomical structure and visualization on the three-dimensional reference framework (**f**).



**Figure 2 | Genome-wide analysis of expression level versus percentage of expressing cells across the entire brain.** **a**, **d**, Histograms of genes versus expression level (**a**, 100 is maximum expression level) or percentage of expressing cells (**d**, 100 is ubiquitous expression). **b**, Box plot indicating median percentage of expressing objects and expression level and distribution across the entire set of expressed genes. Expression scores are normalized to [0,100] with median and third quartiles indicated. **c**, Expression level and percentage of expressing cells are highly correlated among expressing genes, as shown by a regression line ( $R^2 = 0.92$ ). Similar plots for individual structures are available in Supplementary Fig. 1.

binned either by anatomical structure or into three-dimensional voxels of arbitrary size (see Supplementary Methods 2 for mathematical descriptions of variables). To allow visual differentiation of signal from noise, ‘heat maps’ representing signal intensity (average pixel intensity/expressing object) are generated for each two-dimensional image. Pseudo-colouring of these maps by signal intensity provides a visually accessible quantification of ISH signal in individual cells within individual sections (Fig. 1e). An interactive application, Brain Explorer, was also developed to allow visualization of three-dimensional representations of gene expression superimposed on the reference atlas (Fig. 1f).

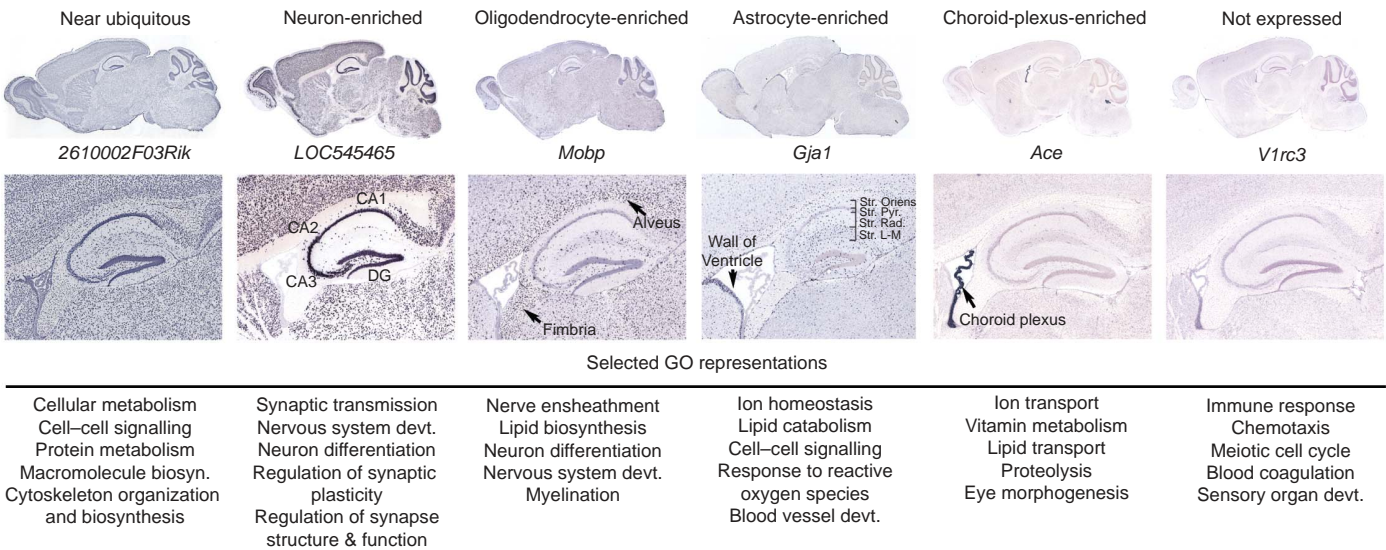
**Expressed versus non-expressed genes.** Approximately 80% of total genes assayed display some cellular expression above background in the brain. This number was equivalently derived through manual scoring and informatics-based analysis using expression thresholds calibrated through expert confirmation. This percentage is significantly higher than predicted from microarray analysis of brain tissues<sup>3</sup>, indicating that many genes are expressed at low levels and/or in small numbers of cells for which expression levels fall below the sensitivity of arrays applied to large brain regions. The distribution of expression level and percentage of expressing cells, and the correlation between these two variables across the entire set of expressers, is plotted in Fig. 2. Notably, most genes are expressed in a relatively small percentage of cells (70.5% of genes are expressed in less than 20% of total cells). Although expression level correlates strongly with the percentage of expressing cells ( $R^2 = 0.92$ ), the distributions of these variables differ from one another (chi-squared,  $P < 2.2 \times 10^{-16}$ ). Similar distributions are observed for individual brain regions (Supplementary Fig. 1). Although expression level can vary as a function of probe length, the distribution is unchanged between the entire set of expressing genes and a subset with probe lengths  $\pm 0.5$  standard deviations (s.d.) from the mean (649–826 nucleotides,  $P = 0.318$ ).

To search for ‘housekeeping’ genes necessary for the basal function of all cells, we examined the genes with the greatest percentage of expressing cells. Although true ubiquity cannot be established without a nuclear counter stain, genes with ‘near ubiquity’ can be validated by visual confirmation of expression in essentially all cells relative to Nissl staining. A total of 186 of the top 500 genes showed detectable expression in all cell classes, although very few genes (for

example, *2610002F03Rik* and *Fbxo22*; Fig. 3) are expressed at high levels in all easily definable cell types (neurons, astrocytes, oligodendrocytes, choroid plexus, ependymal lining of ventricles and pial surface). A much larger set of genes appears to be expressed in all cells but at very low levels in some cellular populations. As expected, gene ontology (GO)<sup>16</sup> categories involving cellular metabolism are highly represented in apparently ubiquitous genes.

**Enriched expression in major cell types.** A targeted search for genes enriched in major cell classes with distinctive anatomical localization is enabled by the atlas grid coordinate system superimposed on each gene’s expression pattern. These coordinates enable searches for genes having similar expression patterns to genes of interest. Seeding with oligodendrocyte-specific genes (*Mbp*, *Mobp*, *Cnp1*) returns a large set of oligodendrocyte-enriched genes; seeding with choroid-plexus-enriched genes (*Col8a2*, *Lbp*, *Msx1*) returns a large set of choroid-enriched genes. Where spatial distributions of cell classes overlap greatly, such as for neurons and astrocytes, this strategy is less robust but can still be effective. Essentially all well-established markers for different cell types were identified with these informatics-based methods for prediction followed by expert confirmation (Supplementary Table 1). GO analysis of genes enriched in major cell classes reflects the different functions of these cells (Fig. 3 and Supplementary Table 2). For example, many oligodendrocyte-enriched genes are involved in myelin production. Certain functional ontologies are over-represented in the non-expressed set as well, including immune responses, meiosis and sensory organ development. Many previously uncharacterized genes are enriched in each cell class, providing novel gene candidates for cell-type-specific function.

**Regionally enriched gene expression.** Genes with regionalized expression patterns provide potential substrates for functional differences between brain regions, whereas correlations between regions may predict functional similarities. To identify the most specific genes for each of 12 major brain regions, the area occupied by signal in each structure was divided by the area occupied by signal in the entire brain for each gene. These ranked lists were visually confirmed to identify the top 100 most specific genes for each region. These lists of enriched genes (Supplementary Table 3) are representative but not comprehensive, as the search missed several well-known genes with high specificity (for example, *Agrp* in the hypothalamus) owing to



**Figure 3 | Representative cell-type-specific genes and corresponding molecular functions.** Examples of genes expressed in all cell types, neurons, oligodendrocytes, astrocytes and the choroid plexus, and a representative non-expressed gene. Selected GO representations<sup>13</sup> are shown for each class (see Supplementary Tables 1 and 2 for complete gene lists and GO terms). *Mobp*, myelin-associated oligodendrocytic basic protein; *Gja1*, gap junction

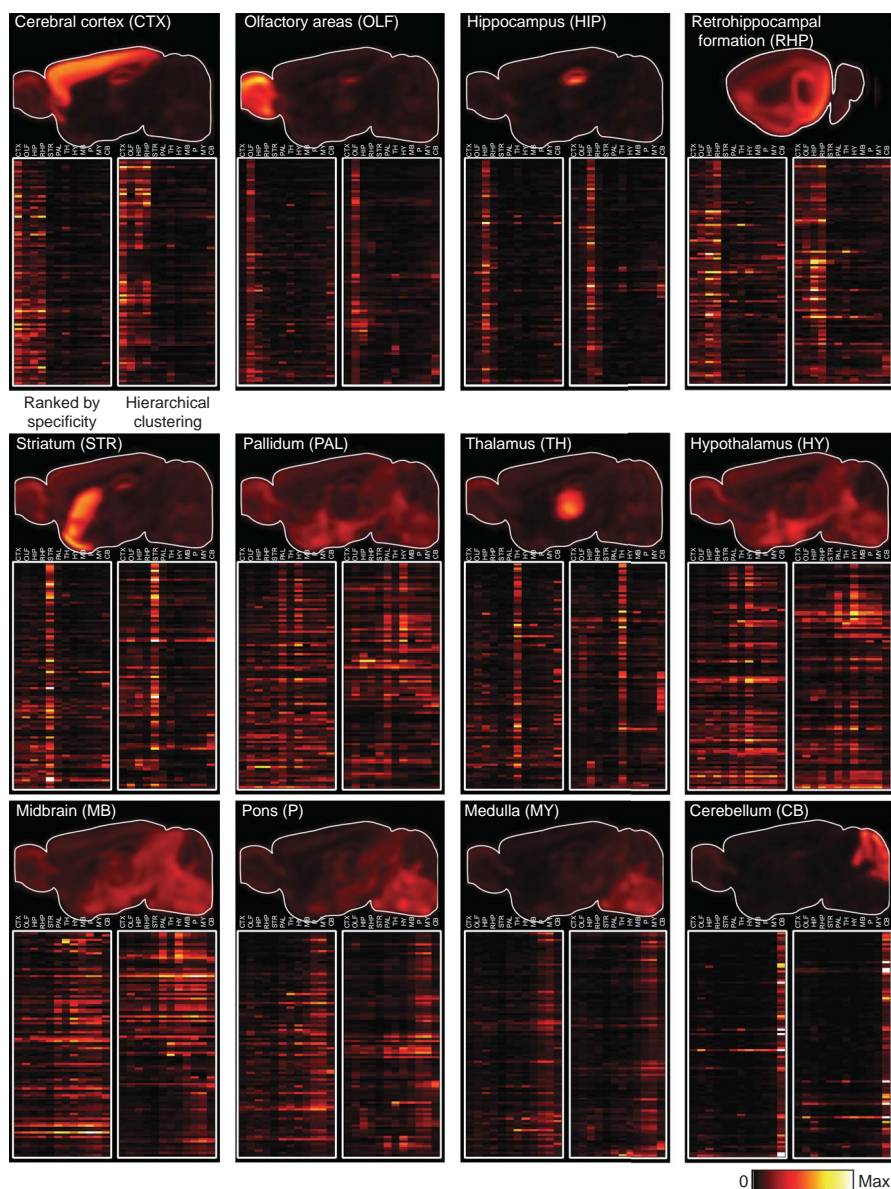
membrane channel protein alpha 1; *Ace*, angiotensin I converting enzyme 1; *V1rc3*, vomeronasal 1 receptor, C3. Hippocampal subfields: CA1, CA2, CA3, DG (dentate gyrus). Layers of CA1: Str. Oriens, stratum oriens; Str. Pyr., stratum pyramidale; Str. Rad., stratum radiatum; Str. L-M, stratum lacunosum-moleculare.



artefacts interfering with accurate registration and/or signal detection. The top candidates for each structure show highly enriched expression, typically restricted to a subset of nuclei/cells within that structure, although few genes are expressed only in a single structure. Some structures have a greater degree of enriched gene expression than others, easily visualized by plotting the average expression of the top 100 genes for a given structure on a representative brain section (Fig. 4). For example, the hippocampus, thalamus, cerebellum and olfactory bulb display a great deal of enriched gene expression. In contrast, the pallidum, hypothalamus, midbrain, pons and medulla exhibit highly overlapping expression patterns.

The complexity of structural combinations is more striking when plotted on a gene-by-gene basis (Fig. 4, bottom row, left column of each panel). Hierarchical clustering of enriched gene sets based on correlations between expression level for each structure reveals unique patterns of correlated gene expression between different structures (Fig. 4, bottom row, right column of each panel). For example, within the cortically enriched gene set, strong correlations are largely with other cortical structures, with relatively distinct gene clusters showing co-expression between combinations of cortical structures. The retrohippocampal formation and hippocampus show strongly correlated expression, as do the pallidum and hypothalamus and the pons and medulla.

**Clustering of correlated gene expression.** Classical definition of brain regions is based on a combination of overall morphology, cellular cytoarchitecture, ontological development and functional connectivity. Grid-level correlations across the brain provide the means to assess whether gene expression in the adult mirrors these organizational principles. To approach this question, correlations were performed between evenly spaced voxels throughout the volume of a brain hemisphere across a set of >5,000 genes chosen to contain a large proportion of genes with regionalized expression but excluding non-expressers or very widely expressed genes. Organizing voxels by inclusion in large anatomical structures (Fig. 5b, d) reveals high correlations between voxels in the same gross anatomical region, and complex but distinct correlations between brain regions. Alternatively, unbiased hierarchical clustering of voxels based on their overall correlations with other voxels produces a different result (Fig. 5c, e). Large anatomical divisions with relatively uniform internal structure, such as the striatum and cerebral cortex, are represented by single tight clusters. Other structures that consist of discrete nuclei, such as the hypothalamus, pons, medulla and midbrain, have local clusters interleaved with one another (Fig. 5e). On the basis of these data, organization of gene expression more closely mirrors the smaller nuclear structure, with stronger relationships between nuclei than between gross anatomical regions.



**Figure 4 | Hierarchical ranking and clustering of the most specific genes in 12 major brain regions.** Top row: average expression of the top 100 enriched genes for each region, plotted as the fractional percentage of expressing pixels/300  $\mu\text{m}^3$  voxel on a representative sagittal section containing that region. Bottom two rows: expression of the same genes across all 12 structures ranked by structural specificity (left column of each panel) or ordered by hierarchical clustering based on correlations between structures (right column of each panel). 'Hot metal' colour scale for percentage of expressing pixels in panels of the top row: dark red (0) to white ( $\geq 0.06$ ); expression level in panels of the bottom two rows: dark red (0) to white ( $> 80$ ).

### Nuclear- and cellular-level analysis of gene expression

A genome-wide ISH data set allows a variety of analyses to understand the fine molecular anatomy of brain regions. One approach is to generate complete expression profiles for gene families of obvious functional relevance, such as the ligand-gated ion channels (Supplementary Data 4 and 5). Alternatively, unbiased analysis of cellular expression patterns in specific brain regions can identify the most specific gene markers for those regions and (putative) cell types, as well as reveal fine, presumably functional subdivisions that are indistinguishable by conventional cytoarchitecture alone.

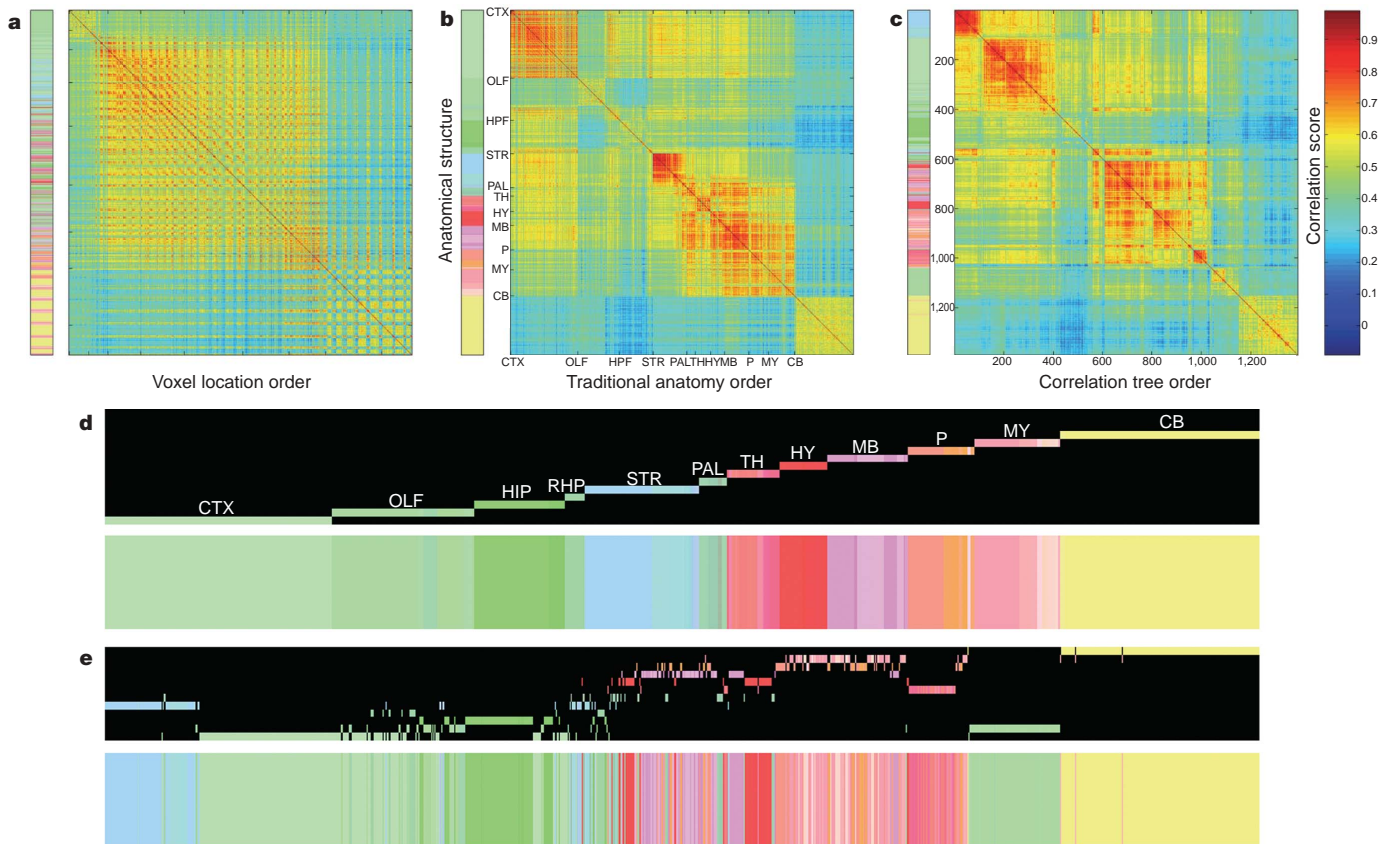
**Laminar and regional cortical specificity.** Systematic analysis of different cortical cell types has been hampered by the lack of specific markers defining neuronal classes with consistent phenotypic properties. Numerous genes have been identified with highly specific cortical gene expression restricted to discrete neuronal populations in different cortical layers, many of which have no known function. For example, *A930038C07Rik* (Fig. 6b) is expressed nearly exclusively in layer 1 cells, potentially Cajal–Retzius cells. Pyramidal neurons in layers 2/3 express the regulator of G-protein signalling family member 8 (*Rgs8*, Fig. 6d), consistent with observations that RGS proteins are involved in higher-level synaptic plasticity, as persistent plasticity is observed in superficial cortical layers<sup>17</sup>. Cocaine-amphetamine regulated transcript (*Cart*; also called *Cartpt*), a neuropeptide involved in reward/reinforcement circuitry<sup>18</sup>, is quite specifically expressed in deep layer 3 of somatosensory cortex (Fig. 6e). Layer 4 is predominantly labelled by *LOC433228* (Fig. 6f). Pyramidal cells in layer 5 are delineated by expression of *C030003D03Rik* (Fig. 6g), whereas layer 6 neurons specifically express the immunoglobulin heavy chain (TIGR accession TC1460681; Fig. 6h). Layer 6b neurons

express the early growth-response gene *Ctgf* (Fig. 6i), as previously described in rat<sup>19</sup>.

Most genes expressed in the neocortex display relatively uniform laminar expression across all cortical areas, consistent with the idea that the basic (canonical) microcircuit is conserved across the entire neocortex. However, functionally discrete cortical regions are also delineated by gene expression, albeit infrequently. For example, the cytochrome P450 family member *Cyp39a1* (Fig. 6j) is expressed almost exclusively in somatosensory cortex, and Marcks-like protein (*Mlp*; also known as *Marcksl1*) is expressed in layer 5 of all neocortical regions except somatosensory cortex (Fig. 6k). These findings suggest that cellular properties vary between different regions of the adult neocortex.

**Heterogeneity in hippocampal subregions.** The hippocampus has been extensively studied owing to its essential role in certain types of learning and memory<sup>20,21</sup>. Lorente de No<sup>22</sup> divided the pyramidal cell layer of the hippocampus proper into CA1, CA2 and CA3 subfields, and there is extensive anatomical evidence for regional specialization within these subfields<sup>22–24</sup>.

As described previously, the main subregions of the hippocampus are specifically delineated by gene expression<sup>25,26</sup>, and a number of additional markers have been identified (Fig. 7a–c). Heterogeneous expression within hippocampal subregions is also common, and is found in all subregions in all cardinal axes. For example, within CA3, *Col15a1* is expressed in a gradient with higher expression proximal to the dentate gyrus (Fig. 7d), and *Crym* is expressed in a reciprocal gradient (Fig. 7e). Heterogeneity is occasionally observed across the depth of the pyramidal cell layer. *Col6a1* is preferentially expressed in a band of cells on the outer border of CA2 and CA3 (Fig. 7f).



**Figure 5 | Correlative analysis of gene expression.** **a**, Correlations between voxels ordered in grid space where colour-coding is based on correlation scores of expression in each voxel to all others across 5,195 genes. **b**, Voxels ordered by structures defined in the Allen Reference Atlas. Colour bar (left) indicates inclusion in listed structures. Most structures display high internal correlations, with differential cross-structural correlations. **c**, Hierarchical

clustering of voxels independent of anatomical location. *x-y* axes in **a–c** represent voxel locations, with the colour scale on the left indicating the anatomical structure in which each voxel is contained. Correlation score bar (range [0,1]) on the right applies to **a–c**. **d, e**, Magnification of colour bars indicating structural identity from **b** and **c**, respectively. Voxel inclusion in the 12 main structures is indicated by a line in the row for each structure.



Dorsal/ventral (septotemporal) heterogeneity is very common (Fig. 7g–m). Individual genes often vary across more than one dimension, defining complex three-dimensional compartments within a given subfield. Five of the ten genes shown in Fig. 7 (selected solely based on specificity) are involved in cell adhesion, and 56 out of 188 cell-adhesion-related genes in the Allen Brain Atlas demonstrate spatially restricted hippocampal expression (data not shown). From a different perspective, 1,137 out of 5,099 analysed genes displayed some form of regionality in the hippocampus. The top over-represented functional category within the regional gene set is cell adhesion ( $P < 1.79 \times 10^{-10}$ ). Differential cell adhesion may be important for establishment and maintenance of topographic connectivity, or, as described recently, different forms of synaptic plasticity and remodelling<sup>27</sup>.

Although there is increasing experimental evidence suggesting functional differentiation along the septotemporal axis of the hippocampus<sup>28–30</sup>, the precise cellular substrate has not been defined. The observation that different neuropeptides (*Grp* and *Nmb*; Fig. 7g–j) are restricted to either dorsal or ventral hippocampus provides evidence for functional differentiation involving discrete signalling pathways. Identification of these regionally restricted markers allows anatomical delineation of novel hippocampal compartments

and provides the means for future experimental manipulation to assess function.

**Novel cerebellar compartments.** The basic structure of the cerebellum is well known and consists of several functionally discrete gross divisions. Additionally, the cerebellar cortex exhibits a bilaterally symmetric series of sagittally oriented bands<sup>31</sup> mirrored by a number of genes, most notably *zebrin*<sup>32,33</sup>. Strong (although not complete) correlation between patterns of cerebellar afferent segregation and *zebrin* expression<sup>34</sup> indicate that molecular markers can delineate functionally discrete regions in the cerebellum.

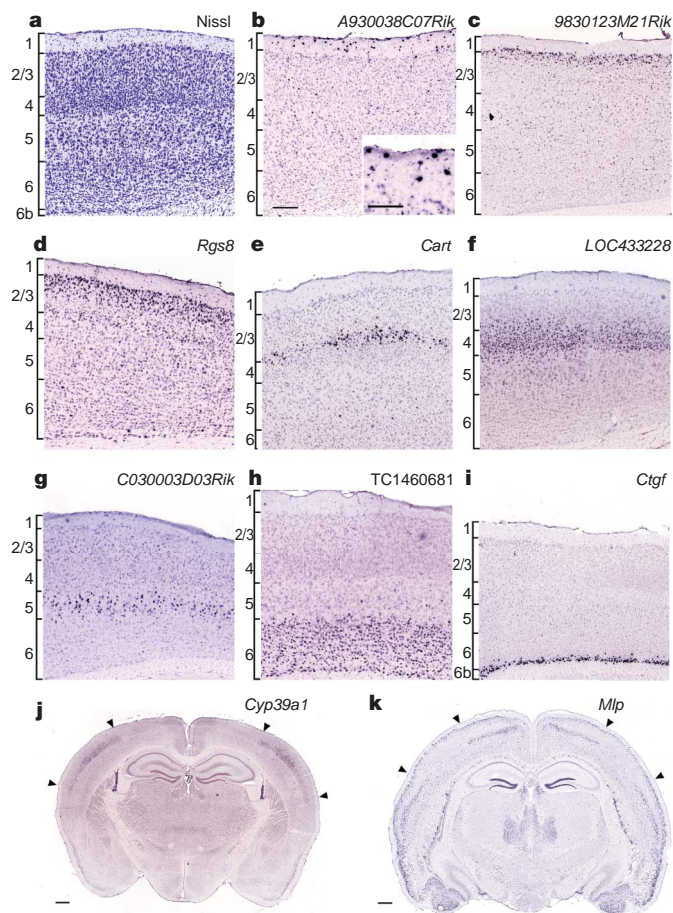
A number of genes display heterogeneity within cerebellar granule and Purkinje cell populations. For example, *Rasgrf1* defines a previously unrecognized large, contiguous domain with sharp boundaries in the granule cell layer of the rostral (Fig. 8a) and dorsal cerebellum (Fig. 8b). More complex regional patterns are observed in the Purkinje cell layer, such as that of *Opn3*, a non-canonical opsin (Fig. 8c, d) whose expression in the cerebellum has been described as a rostro-caudal gradient with radial stripes<sup>35</sup>. Rather than a gradient, three-dimensional reconstruction of ISH data for *Opn3* reveals a more coherent pattern (Fig. 8e) involving a sharply delineated diagonal band lacking expression extending across the entire cerebellum. The overall pattern of *Opn3* is both complex and discrete, with regionalized expression in distinct lobules and sagittal banding in the posterior vermis.

**Subcellular mRNA targeting.** Subcellular localization and translation of mRNA transcripts in dendrites is increasingly recognized as a widespread phenomenon<sup>36</sup>, and is thought to be involved in certain forms of synaptic plasticity<sup>37,38</sup>. Dendritic mRNA targeting is particularly obvious in the hippocampus (Fig. 9a–f) and cerebellum (Fig. 9g, h), where clear distinctions can be made between cell-dense layers, dendritic molecular layers and white matter. Targeting throughout the entire dendritic field is exemplified by the well-characterized patterns of *Camk2a* and *Dnd1* in the hippocampus (Fig. 9b, c)<sup>39</sup>. Although often subtle, this distribution is independent of expression level, thereby distinguishing targeting from passive diffusion of mRNA into dendrites. For example, labelling for the highly expressed gene *Nptx1* is confined to the soma of CA3 pyramidal cells (Fig. 9d), whereas microtubule associated proteins 1A (*Mtap1a*) and 2 (*Mtap2*) also label proximal or proximal and distal dendrites (Fig. 9e, f). Similar forms of targeting are seen in cerebellar Purkinje cells (Fig. 9h, i), as well as in oligodendrocytes (*Mbp*; Fig. 9k)<sup>40</sup>.

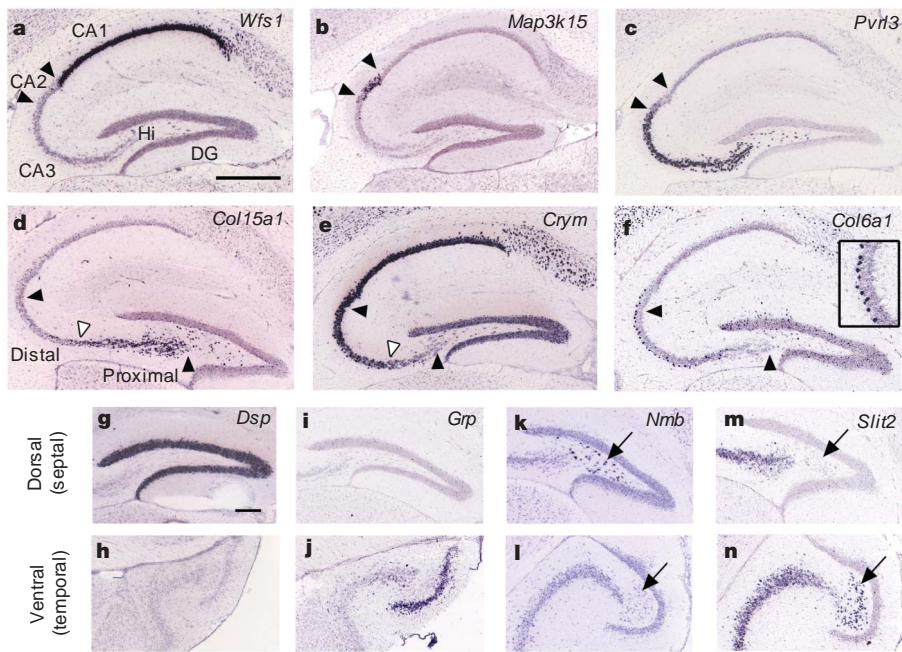
Many genes exhibiting dendritic targeting are involved in cytoskeletal organization and biogenesis, as well as in regulating synaptic plasticity. There appear to be multiple *cis*-acting sequence elements that mediate mRNA targeting<sup>41</sup>. Identification of sets of dendritically targeted mRNAs with shared features (Supplementary Table 4), such as regional specificity and distribution in dendrites, may aid in the identification of conserved sequence elements that correlate with cellular and intracellular transport specificity.

## Discussion

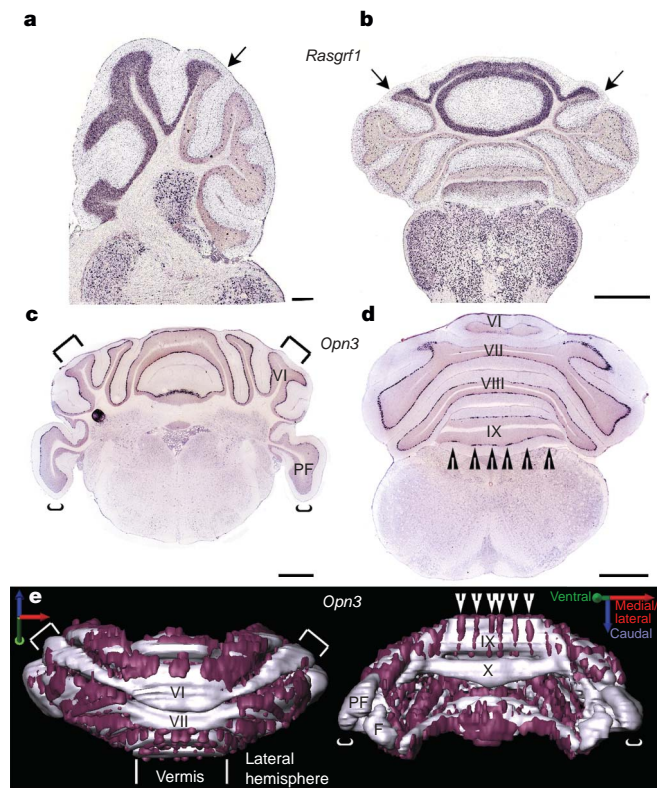
A global analysis of the genetic underpinnings of the brain's structural and cellular complexity requires a genome-wide, cellular resolution, anatomically comprehensive data set. The Allen Brain Atlas has taken a genomics-style approach to understanding this complexity by creating an integrated set of data production and analysis methodologies to systematically produce and analyse a comprehensive atlas of gene expression in the adult C57BL/6J mouse brain. Accurate and comprehensive analysis and annotation of data from the Allen Brain Atlas and other similar projects are difficult but essential hurdles to realizing the full potential of these data. New suites of image analysis tools are required to apply analysis methods developed by the genomics community to cellular-resolution gene expression data in the brain. Ongoing efforts aim to allow correlative cross-gene and cross-structural analysis across the entire Allen Brain Atlas data set by improving automated and semi-automated



**Figure 6 | Laminar and region-specific neocortical gene expression.** **a**, Nissl-stained section with cortical layer boundaries. **b–i**, Layer-specific expression in sagittal sections: *A930038C07Rik* in layer 1 (**b**, inset shows higher magnification view); *9830123M21Rik* in layer 2 (**c**); *Rgs8* in layer 2/3 (**d**); *Cart* in deep layer 3 (**e**); *LOC433228* in layer 4 (**f**); *C030003D03Rik* in layer 5 (**g**); *TC1460681/IgM* in layer 6 (**h**); *Ctgf* in layer 6b (**i**). **j, k**, Selective *Cyp39a1* expression in layer 4 (**j**) and lack of *Mlp* expression in layer 5 (**k**) delineate somatosensory cortex (arrowheads). Scale bars: **b–i**, 200  $\mu$ m; inset in **b**, 50  $\mu$ m; **j, k**, 500  $\mu$ m.



**Figure 7 | Heterogeneity of hippocampal gene expression.** Classically defined hippocampal subregions (black arrowheads) are delineated by gene expression: **a**, *Wfs1* in CA1; **b**, *Map3k15* in CA2; **c**, *Pvrl3* in CA3. Heterogeneous expression within hippocampal subfields: proximal–distal (*Col15a1*, **d**) and distal–proximal (*Crym*, **e**) gradients in CA3 (white arrowheads mark approximate expression boundaries); selective expression of *Col6a1* in the outer band of cells in CA3 (**f**); differential expression in dorsal (*Dsp*, **g**, **h**) and ventral dentate gyrus granule cells (*Grp*, **i**, **j**), and dorsal (*Nmb*, **k**, **l**) and ventral hilar neurons (*Slit2*, **m**, **n**). Scale bars: **a–f**, 500 μm; **g–n**, 200 μm. Hi, hilus (arrows in **k–n**).



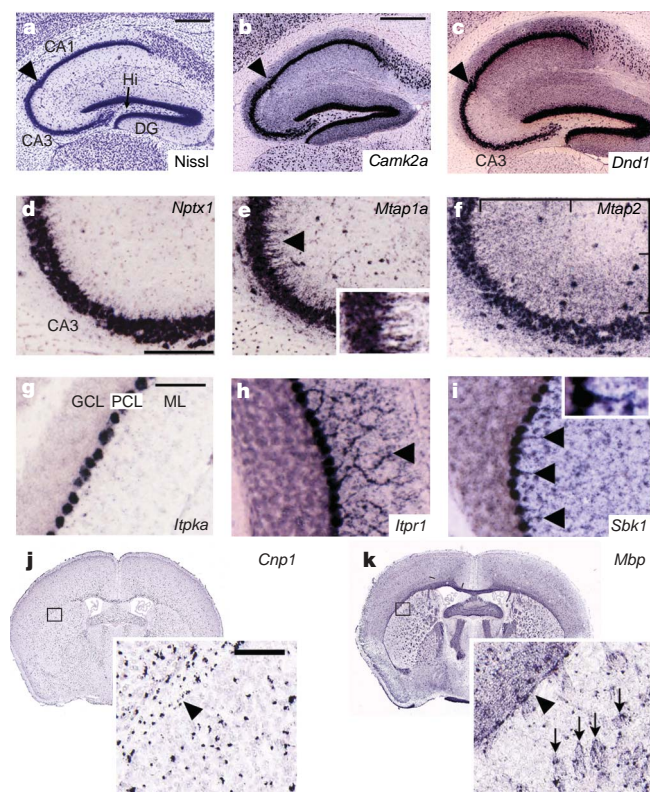
**Figure 8 | Cerebellar compartments revealed by gene expression.** **a**, **b**, Sagittal and coronal views of regional *Rasgrf1* expression in the rostral and dorsal granule cell layer (arrows). **c–e**, Complex regional *Opn3* expression in Purkinje cells in anterior (**c**) and posterior (**d**) sections. **e**, Three-dimensional reconstruction of *Opn3* expression reveals sagittal and diagonal bands. Surface rendering of *Opn3* expression (purple) is superimposed on a model of the granule cell layer (silver) generated from the same brain. Arrowheads mark sagittal bands of *Opn3* expression, whereas brackets distinguish areas lacking *Opn3* expression. Scale bars: **a**, 200 μm; **b–d**, 1 mm. F, flocculus; PF, parafoolliculus.

methodologies in order to quantify and map gene expression at increasingly finer anatomical resolution.

It should be noted that the expression data produced by the current project provide an incomplete snapshot of the molecular complexity of the brain. There is increasing evidence of complex, often cell-type-specific transcriptional splice variation that significantly increases the complexity and size of the transcriptome<sup>42</sup>, whereas a significant proportion of the transcriptome is also dynamically regulated, for example by circadian rhythms<sup>43</sup>. It will also be important to complement these transcriptional data with protein data in the future. In these respects the Allen Brain Atlas serves as a comprehensive baseline data set with which to compare gene expression in other tissues, species, developmental time points, behavioural and disease states.

The identification of large numbers of genes with restricted expression in specific cell populations yields insights into the fine structural and functional organization of specific brain regions and provides the basis for precise functional manipulations. Although soma location and cell density are strong predictors of cell class, important future goals will be to determine the relationship between transcriptional profiles and specific cellular phenotypes by correlating gene expression with more salient characteristics such as morphology, connectivity and physiology<sup>8</sup>. Cell-type-specific transgenic manipulation using promoters derived from highly specific genes is an obvious next step towards these goals, which will be facilitated by systematic promoter analysis of co-expressed genes to identify the necessary *cis*-acting regulatory elements<sup>44</sup> driving gene expression in particular cell types. This powerful approach towards cellular characterization has demonstrated that promoter-driven transgenes can define cell ‘classes’ with consistent morphology and patterns of connectivity<sup>6,11</sup>. Advances in transgenic methodologies to produce spatially and temporally regulated transgene expression promise to revolutionize modern neuroscience, allowing a wide variety of cell-type-specific manipulations including lesions<sup>45</sup>, silencing of neuronal activity<sup>46</sup>, trans-synaptic labelling<sup>47,48</sup> and real-time imaging of neuronal morphology and activity<sup>49</sup>. Region- and cell-type-specific markers identified using the Allen Brain Atlas should broadly facilitate these approaches, allowing specific manipulations to reveal the contributions of specific cell types to higher-level brain function and disease-related dysfunction.





**Figure 9 | Subcellular mRNA targeting.** **a**, Nissl-stained section through the hippocampus. **b**, **c**, Dendritic targeting in apical and basal dendritic zones of CA1–CA3 and DG (*Camk2a*, **b**) or just CA1 and DG (*Dnd1*, **c**). **d–f**, Differential CA3 mRNA targeting: soma only (*Nptx1*, **d**), proximal apical dendrite only (*Mtap1a*, **e**), and proximal and distal dendrites (*Mtap2*, **f**). **g–i**, Cerebellar Purkinje cell targeting: soma only (*Itpka*, **g**), entire dendritic field (*Itp1*, **h**) or proximal dendrites only (*Sbk1*, **i**). **j**, **k**, Oligodendrocyte targeting: soma only (*Cnp1*, **j**) versus white matter (myelin) labelling in external capsule (arrowhead) and striatum (arrows) (*Mbp*, **k**). Scale bars: **a–c**, **g–i**, 200  $\mu$ m; **d–f**, 600  $\mu$ m; **j**, **k**, 100  $\mu$ m. GCL, granule cell layer; ML, molecular layer; PCL, Purkinje cell layer.

## METHODS

High-throughput data generation was achieved by industrializing and automating ISH data production, as well as subsequent image capture for analysis and web-based viewing. This involved creating streamlined pipelines for gene selection, probe manufacturing, tissue sectioning, ISH and *in silico* image processing, and data analysis<sup>50</sup>. Brief descriptions are presented below, and extensive methodological details are supplied as Supplementary Information (Supplementary Methods 1 and 2).

**Animals and tissue processing.** Male, 56-day-old C57BL/6J mice (Jackson Labs West) were maintained on a 12 h light/dark cycle. Brain tissue was collected between 3 and 6 h after light onset. Serial 25- $\mu$ m fresh-frozen cryostat sections were systematically collected starting at a standardized plane of section to ensure reproducible anatomical coverage.

**Riboprobe generation.** A semi-automated process was used to design gene-specific probes based on sequences from multiple sources including Refseq, MGC, Celera, TIGR, Riken and Unigene. Preliminary sequence data were obtained from The Institute for Genomic Research (<http://www.tigr.org>). Probes were designed against unique regions of transcripts to avoid cross-reactivity.

**In situ hybridization.** The semi-automated non-isotopic digoxigenin (DIG)-based ISH platform of ref. 12 was used with minor modifications. Many controls were performed to establish the reproducibility of the platform and to demonstrate both qualitative and quantitative concordance to conventional radioactive ISH methodologies (see Supplementary Data 1–3). Each riboprobe was initially processed on 200- $\mu$ m spaced sagittal sections spanning an entire hemisphere, and a coronal replicate was generated for a subset (~3,500) of genes with restricted expression patterns. An automated image capture platform was developed to digitize ISH data.

**ISH anatomic mapping and quantification.** Mapping of ISH data to a standard coordinate frame was achieved through three-dimensional registration of ISH

data to an annotated volume generated from a Nissl-stained reference atlas created for this project (Supplementary Methods 2 and 3). Rigid registration methods<sup>51</sup> were used to construct the three-dimensional reference atlas, using a 1.5-T, low-resolution, three-dimensional averaged magnetic resonance imaging (MRI) volume as a template to ensure that reconstruction resulted in a realistic volume. Registration of each ISH data set to the Nissl volume involves iterative application of global rigid and subsequently deformable methods<sup>52</sup>.

Signal detection algorithms were developed to identify labelled pixels and segment expressing cells using methods of adaptive image morphology<sup>53</sup> to recognize objects of the correct shape and intensity distribution while ignoring tissue artefacts. The expression mask for each ISH section is constructed by combining tissue boundary detection with small and isolated object recognition (dense or compact nuclei require special treatment). The number of labelled cells proved to be the best predictor of whether or not a gene is expressed in the brain. A minimum cutoff of >1.5% of all 8×8 pixel regions containing some segmented cell body was used to ascertain brain expression, a threshold derived through extensive manual calibration. The two most salient variables for quantitative analysis proved to be relative density (*D*), defined as the proportion of labelled cells, and expression level (*L*), a measure correlated with total transcript count incorporating both area occupied by expressing pixels as well as pixel intensity.

**Global correlation analysis.** Cell-type-enriched markers were identified by searching for genes with highly correlated expression to specific markers for those cell types. A Pearson correlation-based metric was used to compare expression levels between seed genes and the entire Allen Brain Atlas data set in 100- $\mu$ m<sup>3</sup> voxels across the entire brain. *Gja1* was used to search for astrocyte-enriched genes, *Col8a2* for the choroid plexus, *Cnp1* for oligodendrocytes, and *Snrpn* for neurons.

To assess global correlations between different brain regions, this same technique was applied to compare expression levels in a pairwise fashion between each of 1,500 uniformly distributed voxels across the brain. Display of these voxel correlations grouped by (1) inclusion in anatomic structures, or (2) hierarchically clustered by expression correlations between voxels, was computed using Matlab (The MathWorks). Details are provided in Supplementary Methods 4.

We searched for genes enriched in a specific structure by dividing expression in that structure (in this case the fraction of expressing pixels) by expression in the brain as a whole. The resulting ranked gene lists were manually curated to identify the 100 most specific genes. The average expression across the gene lists for each structure was calculated by computing the average expression level *L* on a voxel-by-voxel basis in three-dimensions throughout the brain, and then plotting these values onto a selected plane of section through that structure. Hierarchical clustering based on (Pearson-based) correlations of *L* across 12 brain regions was performed using Matlab.

**Three-dimensional reconstruction of cerebellar gene expression.** Three-dimensional reconstruction of *Opn3* expression was performed using Adobe Photoshop (Adobe Systems) and the 3D Constructor module of Image-Pro Plus (v.5.0; Media Cybernetics) to align and surface render gene expression from uniformly spaced ISH and Nissl images, as described previously<sup>54</sup>.

Received 28 July; accepted 15 November 2006.

Published online 6 December 2006.

1. Zapala, M. A. et al. Adult mouse brain gene expression patterns bear an embryonic imprint. *Proc. Natl Acad. Sci. USA* **102**, 10357–10362 (2005).
2. Datson, N. A., van der Perk, J., de Kloet, E. R. & Vreugdenhil, E. Expression profile of 30,000 genes in rat hippocampus using SAGE. *Hippocampus* **11**, 430–444 (2001).
3. Sandberg, R. et al. Regional and strain-specific gene expression mapping in the adult mouse brain. *Proc. Natl Acad. Sci. USA* **97**, 11038–11043 (2000).
4. Siddiqui, A. S. et al. A mouse atlas of gene expression: large-scale digital gene-expression profiles from precisely defined developing C57BL/6J mouse tissues and cells. *Proc. Natl Acad. Sci. USA* **102**, 18485–18490 (2005).
5. Kamme, F. et al. Single-cell microarray analysis in hippocampus CA1: demonstration and validation of cellular heterogeneity. *J. Neurosci.* **23**, 3607–3615 (2003).
6. Sugino, K. et al. Molecular taxonomy of major neuronal classes in the adult mouse forebrain. *Nature Neurosci.* **9**, 99–107 (2006).
7. Markram, H. et al. Interneurons of the neocortical inhibitory system. *Nature Rev. Neurosci.* **5**, 793–807 (2004).
8. Toledo-Rodriguez, M., Goodman, P., Illic, M., Wu, C. & Markram, H. Neuropeptide and calcium-binding protein gene expression profiles predict neuronal anatomical type in the juvenile rat. *J. Physiol. (Lond.)* **567**, 401–413 (2005).
9. Monyer, H. & Markram, H. Interneuron diversity series: Molecular and genetic tools to study GABAergic interneuron diversity and function. *Trends Neurosci.* **27**, 90–97 (2004).
10. Christiansen, J. H. et al. EMAGE: a spatial database of gene expression patterns during mouse embryo development. *Nucleic Acids Res.* **34**, D637–D641 (2006).

11. Gong, S. *et al.* A gene expression atlas of the central nervous system based on bacterial artificial chromosomes. *Nature* **425**, 917–925 (2003).
12. Visel, A., Thaller, C. & Eichele, G. GenePaint.org: an atlas of gene expression patterns in the mouse embryo. *Nucleic Acids Res.* **32**, D552–D556 (2004).
13. Gray, P. A. *et al.* Mouse brain organization revealed through direct genome-scale TF expression analysis. *Science* **306**, 2255–2257 (2004).
14. Magdaleno, S. *et al.* BGEM: an *in situ* hybridization database of gene expression in the embryonic and adult mouse nervous system. *PLoS Biol.* **4**, e86 (2006).
15. Boguski, M. S. & Jones, A. R. Neurogenomics: at the intersection of neurobiology and genome sciences. *Nature Neurosci.* **7**, 429–433 (2004).
16. Dennis, G. Jr *et al.* DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol.* **4**, P3 (2003).
17. Kirkwood, A., Rioult, M. C. & Bear, M. F. Experience-dependent modification of synaptic plasticity in visual cortex. *Nature* **381**, 526–528 (1996).
18. Elias, C. F. *et al.* Characterization of CART neurons in the rat and human hypothalamus. *J. Comp. Neurol.* **432**, 1–19 (2001).
19. Heuer, H. *et al.* Connective tissue growth factor: a novel marker of layer VII neurons in the rat cerebral cortex. *Neuroscience* **119**, 43–52 (2003).
20. Zola-Morgan, S., Squire, L. R. & Amaral, D. G. Human amnesia and the medial temporal region: enduring memory impairment following a bilateral lesion limited to field CA1 of the hippocampus. *J. Neurosci.* **6**, 2950–2967 (1986).
21. Scoville, W. B. & Milner, B. Loss of recent memory after bilateral hippocampal lesions. 1957. *J. Neuropsychiatry Clin. Neurosci.* **12**, 103–113 (2000).
22. Lorente de Nó, R. Studies on the structure of the cerebral cortex. II. Continuation of the study of the ammonic system. *J. Psychol. Neurol. (Lpz.)* **46**, 113–177 (1934).
23. Amaral, D. G. & Witter, M. P. The three-dimensional organization of the hippocampal formation: a review of anatomical data. *Neuroscience* **31**, 571–591 (1989).
24. Ishizuka, N., Weber, J. & Amaral, D. G. Organization of intrahippocampal projections originating from CA3 pyramidal cells in the rat. *J. Comp. Neurol.* **295**, 580–623 (1990).
25. Tole, S., Christian, C. & Grove, E. A. Early specification and autonomous development of cortical fields in the mouse hippocampus. *Development* **124**, 4959–4970 (1997).
26. Lein, E. S., Zhao, X. & Gage, F. H. Defining a molecular atlas of the hippocampus using DNA microarrays and high-throughput *in situ* hybridization. *J. Neurosci.* **24**, 3879–3889 (2004).
27. Yamagata, M., Sanes, J. R. & Weiner, J. A. Synaptic adhesion molecules. *Curr. Opin. Cell Biol.* **15**, 621–632 (2003).
28. Small, S. A. The longitudinal axis of the hippocampal formation: its anatomy, circuitry, and role in cognitive function. *Rev. Neurosci.* **13**, 183–194 (2002).
29. Moser, M. B. & Moser, E. I. Functional differentiation in the hippocampus. *Hippocampus* **8**, 608–619 (1998).
30. Bannerman, D. M. *et al.* Regional dissociations within the hippocampus—memory and anxiety. *Neurosci. Biobehav. Rev.* **28**, 273–283 (2004).
31. Voogd, J., Hess, D. & Marani, E. *The Parasagittal Zonation of the Cerebellar Cortex in Cat and Monkey: Topography, Distribution of Acetylcholinesterase, and Development* (ed. King, E. S.) (Liss, New York, 1987).
32. Herrup, K. & Kuemerle, B. The compartmentalization of the cerebellum. *Annu. Rev. Neurosci.* **20**, 61–90 (1997).
33. Hawkes, R. & Herrup, K. Aldolase C/zebrin II and the regionalization of the cerebellum. *J. Mol. Neurosci.* **6**, 147–158 (1995).
34. Gravel, C. & Hawkes, R. Parasagittal organization of the rat cerebellar cortex: direct comparison of Purkinje cell compartments and the organization of the spinocerebellar projection. *J. Comp. Neurol.* **291**, 79–102 (1990).
35. Blackshaw, S. & Snyder, S. H. Enkephalin: a novel mammalian extraretinal opsin discretely localized in the brain. *J. Neurosci.* **19**, 3681–3690 (1999).
36. Eberwine, J., Belt, B., Kacharmina, J. E. & Miyashiro, K. Analysis of subcellularly localized mRNAs using *in situ* hybridization, mRNA amplification, and expression profiling. *Neurochem. Res.* **27**, 1065–1077 (2002).
37. Kang, H. & Schuman, E. M. A requirement for local protein synthesis in neurotrophin-induced hippocampal synaptic plasticity. *Science* **273**, 1402–1406 (1996).
38. Huber, K. M., Kayser, M. S. & Bear, M. F. Role for rapid dendritic protein synthesis in hippocampal mGluR-dependent long-term depression. *Science* **288**, 1254–1257 (2000).
39. Burgin, K. E. *et al.* *In situ* hybridization histochemistry of Ca<sup>2+</sup>/calmodulin-dependent protein kinase in developing rat brain. *J. Neurosci.* **10**, 1788–1798 (1990).
40. Trapp, B. D. *et al.* Spatial segregation of mRNA encoding myelin-specific proteins. *Proc. Natl Acad. Sci. USA* **84**, 7773–7777 (1987).
41. Kindler, S., Wang, H., Richter, D. & Tiedge, H. RNA transport and local control of translation. *Annu. Rev. Cell Dev. Biol.* **21**, 223–245 (2005).
42. Carninci, P. *et al.* The transcriptional landscape of the mammalian genome. *Science* **309**, 1559–1563 (2005).
43. Panda, S. *et al.* Coordinated transcription of key pathways in the mouse by the circadian clock. *Cell* **109**, 307–320 (2002).
44. Vavouri, T. & Elgar, G. Prediction of cis-regulatory elements using binding site matrices—the successes, the failures and the reasons for both. *Curr. Opin. Genet. Dev.* **15**, 395–402 (2005).
45. Luquet, S., Perez, F. A., Hnasko, T. S. & Palmiter, R. D. NPY/AgRP neurons are essential for feeding in adult mice but can be ablated in neonates. *Science* **310**, 683–685 (2005).
46. Lechner, H. A., Lein, E. S. & Callaway, E. M. A genetic method for selective and quickly reversible silencing of mammalian neurons. *J. Neurosci.* **22**, 5287–5290 (2002).
47. Yoshihara, Y. *et al.* A genetic approach to visualization of multisynaptic neural pathways using plant lectin transgene. *Neuron* **22**, 33–41 (1999).
48. DeFalco, J. *et al.* Virus-assisted mapping of neural inputs to a feeding center in the hypothalamus. *Science* **291**, 2608–2613 (2001).
49. Brecht, M. *et al.* Novel approaches to monitor and manipulate single neurons *in vivo*. *J. Neurosci.* **24**, 9223–9227 (2004).
50. Ng, L. *et al.* Neuroinformatics for genome-wide 3-D gene expression mapping in the mouse brain. *IEEE Trans. Comput. Biol. Bioinform.* (in the press).
51. Thompson, P. M. & Toga, A. W. in *Handbook of Medical Imaging: Processing and Analysis* (ed. Bankman, I. N.) (Academic Press, San Diego, 2000).
52. Viola, P. & Wells, W. M. Alignment by maximization of mutual information. *Int. J. Comput. Vis.* **24**, 137–154 (1997).
53. Chan, T. F. & Shen, J. *Image Processing and Analysis: Variation, PDE, Wavelet, and Stochastic Methods* (Society for Industrial and Applied Mathematics, Philadelphia, 2005).
54. Lein, E. S., Callaway, E. M., Albright, T. D. & Gage, F. H. Redefining the boundaries of the hippocampal CA2 subfield in the mouse using gene expression and 3-dimensional reconstruction. *J. Comp. Neurol.* **485**, 1–10 (2005).
55. Swanson, L. W. *Brain Maps: Structure of the Rat Brain* (Elsevier, Amsterdam, 2004).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** This work was sponsored by the Allen Institute for Brain Science. The authors wish to thank the Allen Institute founders, P. G. Allen and J. Patton, for their vision, encouragement and support. We also wish to thank key Institute advisors, K. Dooley and S. Coliton, as well as the Scientific Advisory Board for the Atlas project, M. Tessier-Lavigne, D. Anderson, C. Dulac, R. Gibbs, S. Paul, G. Schuler, A. W. Toga and J. Takahashi, for their scientific guidance and dedication to the successful execution of the Atlas project. We would particularly like to acknowledge D. Anderson for his role in the conceptual genesis and continual refinement of Atlas goals, as well as M. Tessier-Lavigne for key scientific and organizational leadership throughout the project. We also thank C. Jennings for his critical reading of the manuscript.

**Author Contributions** Neuroscience Group: E.S.L. (group leader), A.B., L.C., M.P.H., M.T.M., C.L.T., T.A.Z. Informatics Group: M.J.H. (group leader), C.L.K., C.L., L.L.N., S.D.P. Production Groups: Allen Institute for Brain Science: P.E.W. (group leader), S.M.S. (group leader), R.A.J. (group leader), M.A., A.F.B., E.J.B., S.D., N.R.D., A.L.D., T.D., E.D., M.J.D., J.G.D., A.J.E., L.K.E., S.R.F., S.N.G., K.J.G., K.R.H., M.R.H., J.M.K., R.H.K., J.H.L., T.A.L., L.T.L., R.J.M., N.F.M., R.N., G.J.O., T.H.P., S.E.P., O.C.P., R.B.P., Z.L.R., H.R.R., S.A.R., J.J.R., N.R.S., K.S., N.V.S., T.S., C.R.S., S.C.S., K.A.S., N.N.S., K.-R.S., L.R.V., R.M.W., C.K.W., V.Y.W., X.F.Y.; Baylor College of Medicine: C.T. (group leader), N.A., L.C. (Li Chen), T.-M.C., A.C., R.F., A.J.L., Y.L., M.J.R., A.T., M.W., M.B.Y., B.Z.; Max Planck Institute: G.E. (group leader), A.V. Technology Group: C.N.D. (group leader), C.D.T. (group leader), A.B. (Amy Bensinger), K.S.B., M.C.C., J.C., B.E.C., T.A.D., B.J.D., T.P.F., C.F. (Cliff Frensky), D.P.J., P.T.K., R.K., A.R.L., K.D.L., J.M., B.I.S., A.J.S., M.S., R.C.Y., B.L.Y. Other: H.-W.D., B.A.F., C.F. and J.J.M., Allen Reference Atlas generation; J.G.H., data annotation; C.C.O., critical review and manuscript preparation; M.S.B., overall project leadership 2003–2004; A.R.J., overall project leadership 2004–present.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to E.S.L. ([edl@alleninstitute.org](mailto:edl@alleninstitute.org)) or A.R.J. ([allanj@alleninstitute.org](mailto:allanj@alleninstitute.org)).

**Disclaimer** The Nature Publishing Group has a business collaboration with the Allen Institute for the creation and maintenance of the Neuroscience gateway (<http://www.brainatlas.org>), but has no role in generating or curating the Allen Brain Atlas database content. As always, *Nature* Editors have been fully independent and solely responsible for the editorial content and peer review of this research article.



# Thymosin $\beta$ 4 induces adult epicardial progenitor mobilization and neovascularization

Nicola Smart<sup>1</sup>, Catherine A. Risebro<sup>1</sup>, Athalie A. D. Melville<sup>1</sup>, Kelvin Moses<sup>2</sup>, Robert J. Schwartz<sup>2</sup>, Kenneth R. Chien<sup>3</sup> & Paul R. Riley<sup>1</sup>

**Cardiac failure has a principal underlying aetiology of ischaemic damage arising from vascular insufficiency. Molecules that regulate collateral growth in the ischaemic heart also regulate coronary vasculature formation during embryogenesis. Here we identify thymosin  $\beta$ 4 (T $\beta$ 4) as essential for all aspects of coronary vessel development in mice, and demonstrate that T $\beta$ 4 stimulates significant outgrowth from quiescent adult epicardial explants, restoring pluripotency and triggering differentiation of fibroblasts, smooth muscle cells and endothelial cells. T $\beta$ 4 knockdown in the heart is accompanied by significant reduction in the pro-angiogenic cleavage product *N*-acetyl-seryl-aspartyl-lysyl-proline (AcSDKP). Although injection of AcSDKP was unable to rescue T $\beta$ 4 mutant hearts, it significantly enhanced endothelial cell differentiation from adult epicardially derived precursor cells. This study identifies T $\beta$ 4 and AcSDKP as potent stimulators of coronary vasculogenesis and angiogenesis, and reveals T $\beta$ 4-induced adult epicardial cells as a viable source of vascular progenitors for continued renewal of regressed vessels at low basal level or sustained neovascularization following cardiac injury.**

T $\beta$ 4 is a G-actin monomer binding protein implicated in reorganization of the actin cytoskeleton, a process fundamentally required for cell migration. We previously identified T $\beta$ 4 as a putative effector of Hand1-mediated cardiac morphogenesis<sup>1</sup>. More recently T $\beta$ 4 was elevated to a potential therapeutic target (<http://www.regenerx.com/wt/home/index>) in terms of its ability to minimize cardiomyocyte loss after myocardial infarction<sup>2</sup>. Here we reveal an essential role for T $\beta$ 4 in regulating all three key stages of cardiac vessel development (vasculogenesis, angiogenesis and arteriogenesis), and identify the adult epicardium as a potential source of vascular progenitors which, when stimulated by T $\beta$ 4, migrate and differentiate into smooth muscle and endothelial cells. The ability of T $\beta$ 4 to promote coronary vascularization both during development and in the adult enhances cardiomyocyte survival and contributes significantly towards T $\beta$ 4-induced cardioprotection.

## Cardiac-specific knockdown of T $\beta$ 4

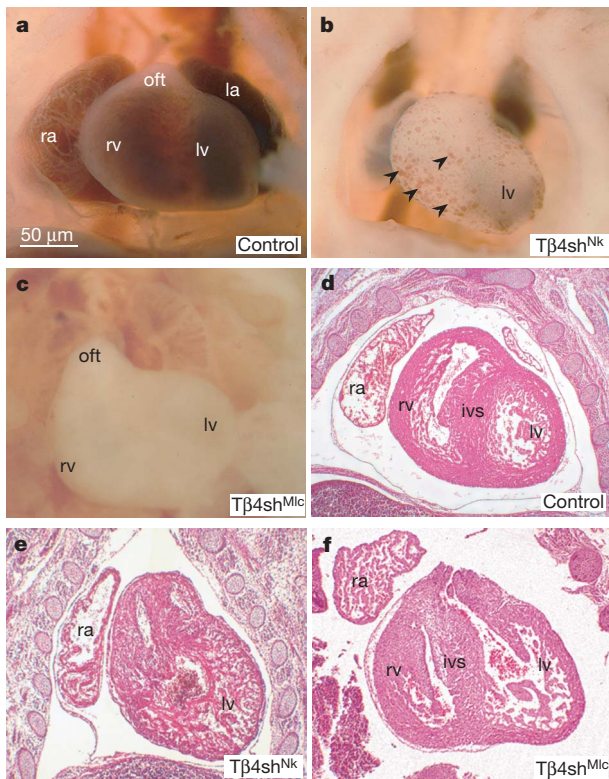
To investigate a role for T $\beta$ 4 during heart development and to provide insight into the mechanism by which the peptide mediates adult cardiac repair, we generated mouse embryos with heart-specific T $\beta$ 4 deficiency using a novel strategy of transgenic conditional RNA interference (RNAi; illustrated in Supplementary Fig. 1a). Floxed T $\beta$ 4 short hairpin RNA (T $\beta$ 4shRNA<sup>flox</sup>) mice were crossed with two strains of Cre-expressing mice: Nkx2-5<sup>Cre</sup>KI (designated T $\beta$ 4sh<sup>NK</sup>), which directs Cre expression throughout the majority of cardiomyocytes<sup>3,4</sup>, and MLC2v<sup>Cre</sup>KI (designated T $\beta$ 4sh<sup>MLC</sup>), which directs Cre expression specifically to ventricular cardiomyocytes<sup>5</sup>. T $\beta$ 4sh<sup>NK</sup> embryos were also observed to have thymic defects (Supplementary Fig. 2a, b) consistent not only with Cre expression driven by Nkx2-5 in the developing thymus<sup>3</sup> (Supplementary Fig. 2c), but also with the thymus representing an obvious source of T $\beta$ 4.

Equivalent phenotypes were observed in embryos from three independent T $\beta$ 4shRNA<sup>flox</sup> transgenic lines. Variation in phenotype between T $\beta$ 4sh<sup>NK</sup> and T $\beta$ 4sh<sup>MLC</sup> embryos could be attributed to variability in the timing and level of Cre-induced recombination (Supplementary Fig. 3a–f). Mutant embryos were categorized into mild, moderate and severe phenotype classes (Supplementary Table 1), and the severity of phenotype correlated directly with the extent of knockdown of T $\beta$ 4 protein (Supplementary Fig. 1b). There was no effect on expression of the closely related  $\beta$ -thymosin, T $\beta$ 10 (Supplementary Fig. 1c), suggesting that the RNAi was specific, although potential off-target effects cannot be unequivocally ruled out at this stage.

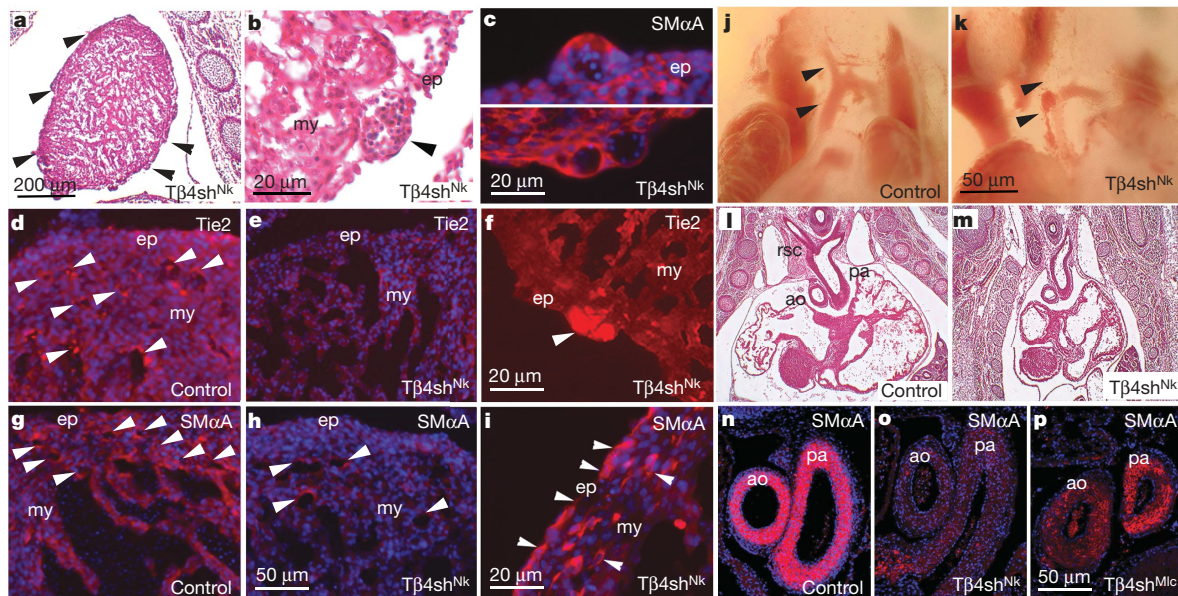
Examination of T $\beta$ 4sh<sup>NK</sup> and T $\beta$ 4sh<sup>MLC</sup> embryos during development revealed no apparent phenotype at embryonic day 8.5 (E8.5), despite reported expression of T $\beta$ 4 in the developing heart from E8.0<sup>1</sup> (data not shown). By E10.5, however, there were clear indications of pericardial oedema in severely affected embryos (Supplementary Fig. 4a–c). At E12.5, T $\beta$ 4sh<sup>NK</sup> mutant embryos had epicardial defects manifesting as abnormal surface nodules and haemorrhaging (Supplementary Fig. 4d,e), before any obvious morphogenetic defects within the myocardium (Supplementary Fig. 4g, h). Severely affected T $\beta$ 4sh<sup>MLC</sup> embryos at E12.5 had, in addition to epicardial defects, severe leftward displacement of the heart with respect to the embryonic midline, and showed early signs of necrosis (Supplementary Fig. 4f).

Detailed analysis of T $\beta$ 4sh<sup>NK</sup> embryos at E14.5 (Fig. 1) revealed a thin non-compacted myocardium and partially detached epicardium mottled with surface nodules (Fig. 1b, e). The ventricular myocardium of T $\beta$ 4sh<sup>MLC</sup> embryos was markedly pale and necrotic, with an extended outflow tract and detached epicardium (Fig. 1c, f). The epicardial defects observed at E12.5 and E14.5 in both T $\beta$ 4sh<sup>NK</sup>

<sup>1</sup>Molecular Medicine Unit, UCL Institute of Child Health, London WC1N 1EH, UK. <sup>2</sup>Department of Molecular and Cellular Biology, Baylor College of Medicine, Houston, Texas 77030, USA. <sup>3</sup>Massachusetts General Hospital Cardiovascular Research Center, Boston, Massachusetts 02114, USA, and the Department of Cell Biology, Harvard Medical School, and the Harvard Stem Cell Institute, Cambridge, Massachusetts 02138, USA.



**Figure 1 | Tβ4 knockdown results in numerous cardiac defects.** Images of whole mount (a–c) and haematoxylin- and eosin-stained frontal sections (d–f) of control and severe Tβ4sh<sup>Nk</sup> and Tβ4sh<sup>Mlc</sup> hearts at E14.5, to demonstrate epicardial nodules indicated by black arrowheads (b), pale, displaced ventricles (c) and failure in compaction of the ventricles (e, f). ra, right atrium; rv, right ventricle; la, left atrium; lv, left ventricle; oft, outflow tract; ivs, interventricular septum.



**Figure 2 | Tβ4 is required for coronary vasculature and great vessel formation.** Frontal sections through the ventricular myocardium (my) of Tβ4sh<sup>Nk</sup> embryos at E14.5, stained with haematoxylin and eosin to visualize epicardial nodules, which represent aberrant coronary vessels; black arrowheads (a, b). SMαA-positive cells surround the cannular epicardial nodules (c). Tie2 immunofluorescence identifies endothelial cells in the control myocardium; white arrowheads (d). Tie2-positive cells are almost absent from Tβ4sh<sup>Nk</sup> myocardium (e) but many appear trapped in epicardial nodules as indicated by white arrowhead (f). Coronary vessels invading the control myocardium are surrounded by SMαA-positive cells; white arrowheads (g). Tβ4sh<sup>Nk</sup> myocardium contains very few vessels and is almost negative for SMαA (h). SMαA-positive cells appear trapped in the epicardium (ep) and compact layer in Tβ4sh<sup>Nk</sup> embryos; white arrowheads (i). Whole mount views (j, k) and frontal sections (l, m) of the developing aorta (ao) and pulmonary artery (pa) of control and Tβ4sh<sup>Nk</sup> embryos at E14.5 show weak vessels and absence of right subclavian artery (rsc; compare black arrowheads indicating position of rsc artery in j compared to k). Absence of SMαA-positive cells in the vessel walls of Tβ4sh<sup>Nk</sup> (o) embryos and a partial loss in Tβ4sh<sup>Mlc</sup> embryos (p), compared with control (n). ep, epicardium; my, myocardium.

and Tβ4sh<sup>Mlc</sup> embryos suggested that coronary vasculogenesis might be perturbed in Tβ4-knockdown mutant hearts<sup>6</sup>. These defects persisted at E15.5 (Supplementary Fig. 4j–l), and those embryos most severely affected were clearly dying at this stage and resorbed by E16.5 (Supplementary Fig. 4o).

### Thymosin β4 is required for coronary vessel development

The epicardial nodules in Tβ4sh<sup>Nk</sup> embryos at E14.5 were cannular, composed of a thin endothelial layer containing a few pericytes, and blood-filled (Fig. 2a–c). Micro-vessels lined with cells positive for the endothelial specific receptor, Tie2 (ref. 7), could be seen invading the dense myocardium of control hearts (Fig. 2d); in contrast, the disrupted myocardium of Tβ4sh<sup>Nk</sup> embryos was almost entirely negative for Tie2, with only a few malformed vessel-like structures apparent (Fig. 2e). Significantly, the epicardial nodules of Tβ4sh<sup>Nk</sup> embryos were intensely stained with Tie2, in clear contrast to the weak expression in the myocardium (Fig. 2f). Coronary vasculogenesis requires that cells delaminate from the epicardium, undergo epithelial-mesenchymal transformation, migrate towards the capillary plexus within the myocardium and differentiate into endothelial cells<sup>6</sup>. The aberrant nodules in the Tβ4sh<sup>Nk</sup> mutants represent a population of Tie2-positive epicardium-derived cells (EPDCs) that have attempted, but failed, to migrate through the myocardium to form coronary vessels. Consistent with impaired vessel development were the thin myocardium and failed ventricular compaction (Figs 1e, 2a, e), a process known to be dependent on epicardially-derived vasculogenesis<sup>8</sup>.

Vascular progenitors in the developing epicardium display bipotency. EPDCs can either form endothelial cells, in response to a combination of myocardial vascular endothelial growth factor (VEGF) and basic fibroblast growth factor (bFGF) signalling, or differentiate into smooth muscle cells, upon exposure to platelet-derived growth factor (PDGF) and transforming growth factor β (TGFβ) (refs 9, 10). Therefore, we next investigated whether there was a reduced incidence of smooth muscle cell recruitment to nascent vessels in Tβ4 knockdown hearts. In controls, smooth muscle cells,



detected by immunostaining for smooth muscle  $\alpha$ -actin (SM $\alpha$ A), were evident throughout the myocardium specifically surrounding the lumen of micro-vessels (Fig. 2g), whereas in T $\beta$ 4sh<sup>Nk</sup> hearts the majority of SM $\alpha$ A-positive cells persisted in the epicardium and in the reduced compact layer, rarely localized to the inner trabeculae (Fig. 2c, h, i). At earlier stages (E12.5), SM $\alpha$ A-positive cells were barely detectable in the myocardium of T $\beta$ 4sh<sup>Nk</sup> mutants and only sparsely distributed in T $\beta$ 4sh<sup>Mlc</sup> mutants compared to controls (Supplementary Fig. 5a). This suggests that, as a consequence of reduced T $\beta$ 4 signalling from the myocardium, there is a significant delay in smooth muscle cell differentiation. Moreover, EPDCs, fated to form smooth muscle cells, fail to migrate extensively into the myocardium to provide support to the coronary vessels and instead activate the smooth muscle cell differentiation program *in situ*.

Defective recruitment of smooth muscle cells, so-called collateral growth<sup>10</sup>, was also observed in the apical media of the large thoracic vessels of T $\beta$ 4sh<sup>Nk</sup> (and to a lesser extent, T $\beta$ 4sh<sup>Mlc</sup>) embryos. Both aorta and pulmonary artery were morphologically abnormal and visibly defective in terms of both angiogenesis and arteriogenesis, with decreased vessel wall stability and haemorrhaging (Fig. 2j–p). Failed contribution of EPDC-derived smooth muscle cells may explain in part the loss of SM $\alpha$ A-positive cells from the mutant thoracic vessels (Fig. 2o, p); further loss probably arises as a consequence of failure of the coronary vasculature to establish appropriate collateral connections to the aortic root<sup>10</sup>.

The defects associated with T $\beta$ 4 knockdown in the myocardium are clearly non cell-autonomous. Myocardial compaction, coronary vasculogenesis and smooth muscle cell recruitment all relate to the epicardial lineage, and specifically to EPDC differentiation and appropriate migration<sup>6,8</sup>. T $\beta$ 4 is expressed throughout the developing myocardium, but is expressed neither in the epicardium (Supplementary Fig. 6) nor in endothelial cells residing specifically within the wall of the heart at equivalent stages (Supplementary Fig. 3g–j). Therefore, knockdown of intracellular T $\beta$ 4 in the myocardium results in loss of functional secreted T $\beta$ 4 and impaired paracrine signalling to the epicardium. Myocardial signalling associated with coronary vessel development from epicardium has also been attributed to the GATA co-factor friend of GATA 2 (FOG-2) (ref. 11) and to VEGF (ref. 12). However, FOG-2 and VEGF are implicated exclusively in the induction of coronary vasculogenesis, whereas T $\beta$ 4 uniquely affects all aspects of coronary vessel development.

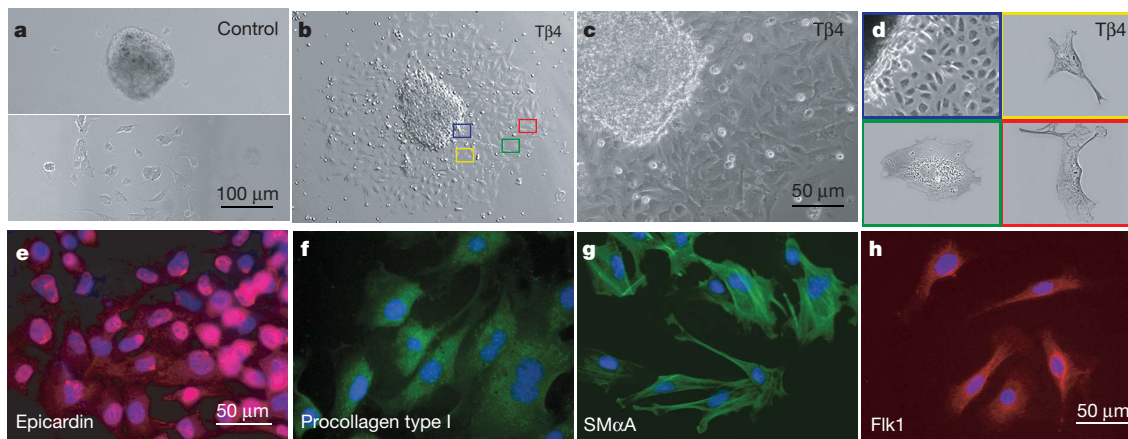
### T $\beta$ 4 promotes vascularization from embryonic epicardium

In order to support our *in vivo* phenotype analyses and assess a direct effect of T $\beta$ 4 on developing epicardium, we established epicardial

explant cultures from wild-type hearts<sup>13</sup>. We initially derived epicardial explants from stages E10.5 to E16.5 and postnatal day 1 (P1) neonates, treated with either T $\beta$ 4, with VEGF and FGF7 (ref. 14), or a combination of T $\beta$ 4 and growth factors. Explants from E10.5 hearts, coincident with the formation of the epicardium, produced extensive outgrowths that differentiated into SM $\alpha$ A- and Tie2-positive cells (Supplementary Fig. 5b). Addition of T $\beta$ 4 significantly increased the numbers of SM $\alpha$ A- and Tie2-positive cells, and these cell populations were enhanced further still with the addition of VEGF and FGF7 (Supplementary Fig. 5b). The most potent effect was observed following combined treatment of T $\beta$ 4 and VEGF/FGF7, which resulted in a highly significant increase in Tie2-positive endothelial cells. Synergistic function of T $\beta$ 4 with VEGF in this assay is consistent with our observed downregulation of VEGF in T $\beta$ 4 knockdown hearts (Supplementary Fig. 5c), suggesting appropriate VEGF expression may require T $\beta$ 4. The ability of the epicardial explants to produce outgrowths of cells competent to differentiate in response to treatment significantly diminished over the course of development as observed at E12.5 (Supplementary Fig. 5b), and continued to decrease such that by E16.5 and P1 neonate stages there was no detectable expression of either SM $\alpha$ A or Tie2 (data not shown). This suggests that isolated epicardial derived cells after emerging from the epicardium have an optimal window of time in development (E10.5–12.5) during which they can respond to myocardial signals and differentiate into vascular progenitors.

### T $\beta$ 4 induces adult epicardial cell migration

As coronary vasculogenesis is required to maintain cardiomyocyte survival and consequently appropriate myocardial architecture and cardiac function, the role of T $\beta$ 4 in coronary vessel development may underlie its reported ability to act therapeutically in terms of cardioprotection and repair<sup>2</sup>. Translation of a vascular development role for T $\beta$ 4 to that of angiogenic therapy for coronary artery disease in the adult heart requires releasing the adult epicardium from a quiescent state and restoring its pluripotency. To investigate the potential for T $\beta$ 4 in this context, we isolated epicardial explants from wild-type adult mouse hearts at 8–12 weeks of age (Fig. 3), and assessed the ability of T $\beta$ 4 to induce outgrowth in addition to any differentiation phenotype. Untreated adult explants demonstrated virtually no detectable outgrowth (Fig. 3a), with only a few isolated cells observed in the culture dish (Fig. 3a). This is consistent with the adult epicardium having lost migration, differentiation and signalling capacities during the latter half of gestation<sup>13</sup>. In contrast, treatment with T $\beta$ 4 stimulated extensive outgrowth of cells that differentiated into a variety of discernable phenotypes (Fig. 3b–d). The emerging epithelial cells



**Figure 3 | T $\beta$ 4 promotes migration of adult EPDCs and enables their differentiation into vasculogenic cells.** Outgrowth of large colonies of cells from adult heart explants stimulated by T $\beta$ 4 (b, c), compared with a minimal degree of migration from untreated explants (a). Emerging cells (b, d, blue box) identified as epicardial cells by immunostaining for epicardin (e).

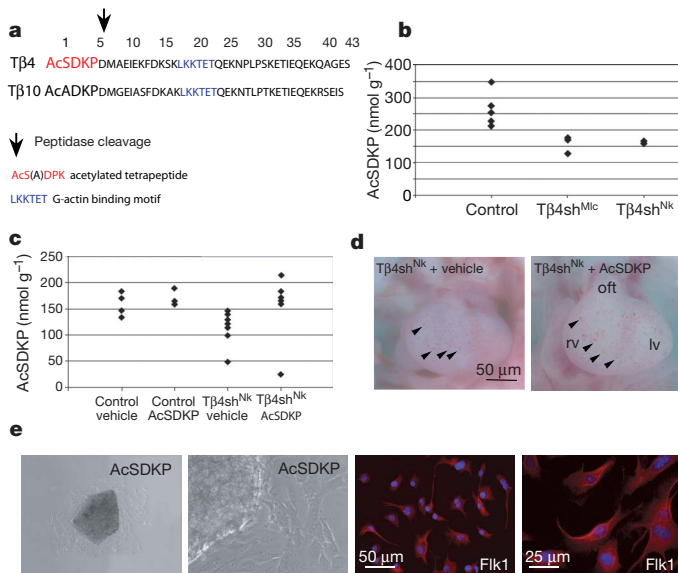
Following migration, cells undergo differentiation into smooth muscle cells, identified by immunostaining for SM $\alpha$ A (b, d green box, g), fibroblasts identified by immunostaining for procollagen type I (b, d yellow box, f) and endothelial cells identified by immunostaining for Flk-1 (b, d red box, h).

were positive for the epicardial-specific transcription factor epicardin (Fig. 3e), and these cells differentiated with migration into procollagen type I-, SM $\alpha$ A- and Flk1 (VEGF receptor)-positive cells indicative of fibroblasts, smooth muscle and endothelial cells (Fig. 3f–h).

EPDCs account for a significant proportion of the 50% of cells in the murine cardiac wall that are not cardiomyocytes<sup>15</sup>; their contribution is assumed to be confined entirely to embryonic epicardium. Remarkably, the explant studies reveal for the first time an equivalent potential within the adult lineage. Vascular regeneration includes adaptive vasculogenesis and arteriogenesis<sup>16</sup>, and the supply of endothelial and smooth muscle vascular precursors required for this process has been attributed, in part, to the peripheral circulation and bone marrow<sup>17,18</sup>. Here we demonstrate the enormous potential of the adult epicardium under the control of T $\beta$ 4. Release of quiescent EPDCs represents a viable source of vascular progenitors for continued renewal of regressed vessels at low basal level or sustained neovascularization following cardiac injury. In ischaemic and failing hearts, T $\beta$ 4-treated adult EPDCs have the potential to deliver endothelial, smooth muscle and fibroblastic cells to sites of injury that need additional sustenance, thus ensuring adequate perfusion of damaged heart muscle and structural integrity of the myocardium.

### T $\beta$ 4 dependent vascularization underlies cardiomyocyte survival

A major consequence of T $\beta$ 4 knockdown in the mutant hearts was a failure to maintain an integrated actin cytoskeleton, as revealed by global disruption of F-actin (Supplementary Fig. 7a–c). A high proportion of cells in the myocardium of both T $\beta$ 4sh<sup>NK</sup> and T $\beta$ 4sh<sup>Mlc</sup> mutant hearts showed Fas clustering at the cell membrane, indicative of Fas activation (Supplementary Fig. 7d–i), increased production of the cleaved (active) forms of initiator caspase-8 and effector caspase-3 consistent with induction of an apoptotic cascade (Supplementary Fig. 7b), and elevated TUNEL (TdT-mediated dUTP nick end

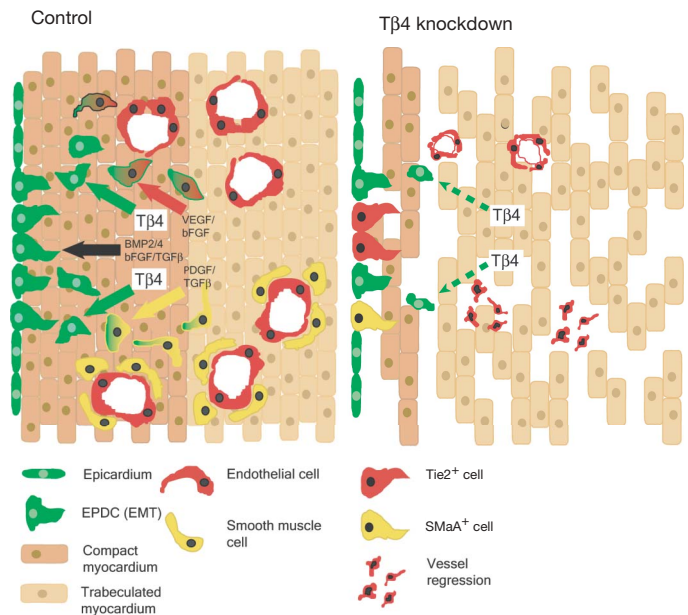


**Figure 4 | AcSDKP is not sufficient to compensate for loss of T $\beta$ 4 in coronary vessel development.** AcSDKP can be produced by peptidase cleavage of T $\beta$ 4 but not T $\beta$ 10, the closely related  $\beta$ -thymosin expressed in the developing heart (a). AcSDKP levels are reduced in T $\beta$ 4sh<sup>Mlc</sup> (60%,  $P < 0.01$ ,  $n = 5$ ) and T $\beta$ 4sh<sup>NK</sup> (62%,  $P < 0.01$ ,  $n = 4$ ) hearts as a result of T $\beta$ 4 knockdown (b). AcSDKP injection into pregnant females restored the peptide to control level in hearts of T $\beta$ 4sh<sup>NK</sup> embryos (c). Restoration of AcSDKP in the developing heart is insufficient to rescue vascular defects associated with knockdown of T $\beta$ 4 (d); T $\beta$ 4sh<sup>NK</sup> hearts with AcSDKP equivalent to control level display characteristic defects of severe T $\beta$ 4sh<sup>NK</sup> knockdown hearts, such as the epicardial nodules highlighted by black arrowheads in (d). AcSDKP promotes the differentiation of adult EPDCs into Flk1-positive endothelial cells (e).

labelling) staining as evidence of increased DNA fragmentation (Supplementary Fig. 7l–n). This supports a role for T $\beta$ 4 in maintenance of cardiomyocyte survival, previously thought to be due to T $\beta$ 4-induced activation of the survival kinase and anti-apoptotic factor, Akt (also known as protein kinase B (PKB) (ref. 2). We confirmed early activation of Akt in C2C12 (mouse myoblast) cells treated with T $\beta$ 4 as previously described<sup>2</sup> (Supplementary Fig. 8a), but contrary to expectation active Akt-P (Ser 473) was not downregulated and was significantly higher in T $\beta$ 4sh<sup>NK</sup> and T $\beta$ 4sh<sup>Mlc</sup> hearts than in control hearts (Supplementary Fig. 8b). From a cell survival perspective this finding suggests that T $\beta$ 4 is unlikely to directly regulate Akt and that Akt activation may be a secondary response to T $\beta$ 4 treatment, or that the increased activation of Akt may be a compensatory mechanism in the absence of T $\beta$ 4.

### AcSDKP stimulates endothelial cell differentiation from adult epicardium

In our model, preservation of myocardium is secondary to T $\beta$ 4-induced coronary vasculogenesis, angiogenesis and collateral growth. The mechanism by which T $\beta$ 4 stimulates coronary vessel development in this regard involves T $\beta$ 4 directly promoting EPDC migration from the epicardium via its previously known function of actin binding, filament assembly and lamellipodia formation. However, scope exists for a non actin-mediated vasculo-, angio- and arteriogenic function for T $\beta$ 4 by virtue of its endoprotease activity to produce the pro-angiogenic tetrapeptide *N*-acetyl-seryl-aspartyl-lysyl-proline (AcSDKP; Fig. 4a)<sup>19–22</sup>. We therefore quantified AcSDKP levels in our mutant knockdown hearts by competitive enzyme immunoassay on extracted myocardium, and found they



**Figure 5 | Model for the role of T $\beta$ 4 in coronary vessel development.**

During normal heart development (left panel), EPDCs undergo an epithelial to mesenchymal transition (EMT) in response to BMP, FGF and TGF $\beta$  signalling from the myocardium. Secreted myocardial T $\beta$ 4 then induces EPDCs to migrate into the myocardium where they respond to angio-(VEGF/bFGF) or arteriogenic (PDGF/TGF $\beta$ ) factors and differentiate into either endothelial or smooth muscle cells respectively, thus establishing a capillary plexus and stabilization of coronary vessels. Following knockdown of myocardial T $\beta$ 4 (right panel), EPDCs undergo EMT but fail to migrate, becoming trapped in the epicardium where they differentiate into Tie2- and SM $\alpha$ A-positive cells. Failure in T $\beta$ 4-induced EPDC migration results in significantly impaired vasculogenesis, defective collateral growth and vascular regression, which in turn leads to a severe reduction in cardiomyocyte survival manifested as a thin compact layer and disrupted myocardial architecture.



were decreased to 62% and 60%, respectively, of that of controls (Fig. 4b,  $P < 0.01$ ); this is robust evidence for a peptide and precursor peptide relationship between T $\beta$ 4 and AcSDKP in a physiological setting.

We next investigated whether AcSDKP could rescue any of the vasculogenesis defects observed in the T $\beta$ 4 mutant hearts. Intraperitoneal injection of AcSDKP into pregnant females successfully restored the peptide to control levels in mutant embryo hearts at E14.5 (Fig. 4c), but failed to rescue any of the phenotype attributed to T $\beta$ 4 knockdown (Fig. 4d). This supports our interpretation of the primary phenotype in T $\beta$ 4 mutants. AcSDKP lacks actin binding function, rendering it unable to stimulate filamentous actin assembly and lamellipodia-based cell migration and consequently unable to rescue the EPDC defect. However, in the adult, consistent with reported cardioprotective effects of AcSDKP (refs 23–25), we observed a significant upregulation in levels of both endogenous T $\beta$ 4 and AcSDKP in response to ischaemia after 1 day and 1 week, respectively (Supplementary Fig. 9). Moreover, addition of AcSDKP to adult epicardial explants resulted in a striking increase in differentiated (Flk1-positive) endothelial cells (Fig. 4e). Although unable to promote epicardial outgrowth beyond control levels, AcSDKP brought about rapid differentiation of any emerging EPDCs, such that Flk1-positive cells were observed in close proximity to the explant tissue (Fig. 4e). The differentiated cells were almost exclusively endothelial, with only very few smooth muscle cells observed in AcSDKP-treated cultures (data not shown). This suggests that cleavage of AcSDKP from T $\beta$ 4 exclusively promotes EPDC endothelial cell differentiation, and may underlie a compound vasculogenic effect of T $\beta$ 4 aside from simply promoting EPDC migration into overlying myocardium as an instructive cue for differentiation. Crucial to the further understanding of this two-step function will be the identification of the respective receptors for T $\beta$ 4 and AcSDKP.

## Overview

T $\beta$ 4 is presented here as a single factor that can potentially couple myocardial and coronary vascular regeneration in failing mouse hearts. A major shortcoming of current angiogenic therapy in response to myocardial ischaemia in humans is that the outcome may be limited to capillary growth without concomitant collateral support of arterioles<sup>26</sup>. Our findings that, in mice, T $\beta$ 4 can promote vessel formation and collateral growth not only during development (Fig. 5) but also critically from adult epicardium, suggest T $\beta$ 4 has considerable therapeutic potential in humans. To this end, we reveal the mechanism by which T $\beta$ 4 may act to promote cardiomyocyte survival following acute myocardial damage in mice, and identify the biopeptide AcSDKP as a small molecule that potentially offers further protection following cardiac injury.

## METHODS

Detailed methods are given in Supplementary Information.

**Construction of T $\beta$ 4 shRNA transgene.** 23-base-pair sense and antisense T $\beta$ 4 sequences were inserted downstream of the H1 RNA pol III promoter in a modified pcDNA3 vector<sup>27</sup>. Sequence of the shRNA: CTGAGATCGAGAAA TTCGATAAG; nucleotides 151–173 (Accession: NM 021278.2 *Mus musculus* thymosin  $\beta$ 4, X chromosome (Tmsb4x) mRNA). A 5T stop sequence, flanked by two loxP recombination sequences, was inserted upstream of the hairpin sequences to enable activation of T $\beta$ 4 knockdown by Cre recombinase. This construct was used to derive transgenic mouse lines.

**AcSDKP enzyme immunoassay.** Total protein from embryonic or adult heart was extracted in methanol, and AcSDKP concentration was measured in triplicate using an enzyme immunoassay kit.

**AcSDKP rescue in T $\beta$ 4 knockdown embryos.** Four pregnant females (T $\beta$ 4 shRNA  $\times$  Nkx2-5 heterozygote cross) were injected intraperitoneally with 36  $\mu$ g AcSDKP and two with vehicle at E6.5, E8.5, E10.5 and E12.5. Embryos were dissected at E14.5 and assessed for severity of phenotype. Western blotting and immunoassay were used to determine levels of T $\beta$ 4 and AcSDKP in hearts.

**Culture of EPDCs from embryonic and adult heart.** Embryonic EPDCs were cultured and passaged as previously described<sup>13</sup>. After first passage, cells were treated with or without T $\beta$ 4 (100 ng ml<sup>-1</sup>) and/or recombinant VEGF and FGF7

(10 ng ml<sup>-1</sup>) for 48 h, before collection of protein extracts. Adult hearts were removed from 8-week-old C57Bl/6 mice, cut into pieces and plated onto gelatin-coated dishes in DMEM containing 15% FBS in the presence or absence of T $\beta$ 4 (100 ng ml<sup>-1</sup>) or AcSDKP (18  $\mu$ g ml<sup>-1</sup>). After 72 h, cells were gently washed and medium replaced; after a further 24 h, cells were fixed and cell types assessed by immunofluorescence.

**Myocardial infarction.** Myocardial infarction was induced in anaesthetized C57Bl/6 male mice by ligation of the left anterior descending coronary artery (permanent occlusion). Animals were killed 1 h, 1 day or 1 week after myocardial infarction, and protein extracts were prepared from heart.

Received 8 August; accepted 25 October 2006.

Published online 15 November 2006.

- Smart, N., Hill, A. A., Cross, J. C. & Riley, P. R. A differential screen for putative targets of the bHLH transcription factor Hand1 in cardiac morphogenesis. *Mech. Dev.* **119**, S65–S71 (2002).
- Bock-Marquette, I., Saxena, A., White, M. D., Dimaio, J. M. & Srivastava, D. Thymosin  $\beta$ 4 activates integrin-linked kinase and promotes cardiac cell migration, survival and cardiac repair. *Nature* **432**, 466–472 (2004).
- Moses, K., DeMayo, F., Braun, R., Reedy, J. & Schwartz, R. Embryonic expression of an Nkx2-5/Cre gene using ROSA26 reporter mice. *Genesis* **31**, 176–180 (2001).
- Zeisberg, E. M. *et al.* Morphogenesis of the right ventricle requires myocardial expression of *Gata4*. *J. Clin. Invest.* **115**, 1522–1531 (2005).
- Chen, J. *et al.* Selective requirement of myosin light chain 2v in embryonic heart function. *J. Biol. Chem.* **273**, 1252–1256 (1998).
- von Kodolitsch, Y. *et al.* Coronary artery anomalies Part I: Recent insights from molecular embryology. *Z. Kardiol.* **93**, 929–937 (2004).
- Ward, N. L. & Dumont, D. J. The angiotensins and Tie2/Tek: adding to the complexity of cardiovascular development. *Semin. Cell Dev. Biol.* **13**, 19–27 (2002).
- Merki, E. *et al.* Epicardial retinoid X receptor  $\alpha$  is required for myocardial growth and coronary artery formation. *Proc. Natl Acad. Sci. USA* **102**, 18455–18460 (2005).
- Yamashita, J. *et al.* Flk1-positive cells derived from embryonic stem cells serve as vascular progenitors. *Nature* **408**, 92–96 (2000).
- Carmeliet, P. Angiogenesis in health and disease. *Nature Med.* **9**, 653–660 (2003).
- Tevosian, S. G. *et al.* FOG-2, a cofactor for GATA transcription factors, is essential for heart morphogenesis and development of coronary vessels from epicardium. *Cell* **101**, 729–739 (2000).
- Giordano, F. J. *et al.* A cardiac myocyte vascular endothelial growth factor paracrine pathway is required to maintain cardiac function. *Proc. Natl Acad. Sci. USA* **98**, 5780–5785 (2001).
- Chen, T. H. *et al.* Epicardial induction of fetal cardiomyocyte proliferation via a retinoic acid-inducible trophic factor. *Dev. Biol.* **250**, 198–207 (2002).
- Morabito, C. J., Dettman, R. W., Kattan, J., Collier, J. M. & Bristow, J. Positive and negative regulation of epicardial-mesenchymal transformation during avian heart development. *Dev. Biol.* **234**, 204–215 (2001).
- Poelmann, R. E., Lie-Venema, H. & Gittenberger-de Groot, A. C. The role of the epicardium and neural crest as extracardiac contributors to coronary vascular development. *Tex. Heart Inst. J.* **29**, 255–261 (2002).
- Luttun, A. & Carmeliet, P. De novo vasculogenesis in the heart. *Cardiovasc. Res.* **58**, 378–389 (2003).
- Kamihata, H. *et al.* Implantation of bone marrow mononuclear cells into ischemic myocardium enhances collateral perfusion and regional function via side supply of angioblasts, angiogenic ligands, and cytokines. *Circulation* **104**, 1046–1052 (2001).
- Saiura, A., Sata, M., Hirata, Y., Nagai, R. & Makuuchi, M. Circulating smooth muscle progenitor cells contribute to atherosclerosis. *Nature Med.* **7**, 382–383 (2001).
- Grillon, C. *et al.* Involvement of thymosin  $\beta$ 4 and endoprotease Asp-N in the biosynthesis of the tetrapeptide AcSerAspLysPro a regulator of the hematopoietic system. *FEBS Lett.* **274**, 30–34 (1990).
- Rieger, K. J. *et al.* Involvement of human plasma angiotensin I-converting enzyme in the degradation of the haemoregulatory peptide N-acetyl-seryl-aspartyl-lysyl-proline. *Biochem. J.* **296**, 373–378 (1993).
- Goldstein, A. L., Hannappel, E. & Kleinman, H. K. Thymosin  $\beta$ 4: actin-sequestering protein moonlights to repair injured tissues. *Trends Mol. Med.* **11**, 421–429 (2005).
- Wang, D. *et al.* N-acetyl-seryl-aspartyl-lysyl-proline stimulates angiogenesis *in vitro* and *in vivo*. *Am. J. Physiol. Heart Circ. Physiol.* **287**, H2099–H2105 (2004).
- Rasoul, S. *et al.* Antifibrotic effect of Ac-SDKP and angiotensin-converting enzyme inhibition in hypertension. *J. Hypertens.* **22**, 593–603 (2004).
- Pokharel, S. *et al.* Increased myocardial collagen content in transgenic rats overexpressing cardiac angiotensin-converting enzyme is related to enhanced breakdown of N-acetyl-ser-asp-lys-pro and increased phosphorylation of Smad2/3. *Circulation* **110**, 3129–3135 (2004).
- Peng, H. *et al.* Angiotensin-converting enzyme inhibitors: A new mechanism of action. *Circulation* **112**, 2436–2445 (2005).
- Tomanek, R., Zheng, W. & Yue, X. Growth factor activation in myocardial vascularization: Therapeutic implications. *Mol. Cell. Biochem.* **264**, 3–11 (2004).

27. Kunath, T. *et al.* Transgenic RNA interference in ES cell-derived embryos recapitulates a genetic null phenotype. *Nature Biotechnol.* **21**, 559–561 (2003).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** This research was supported by the British Heart Foundation and the Medical Research Council. We acknowledge the collaboration of K. Thiam and S. Millet of genOway (France) in generating the T $\beta$ 4shRNAflox line and G. Gish for providing RasGAPshRNA plasmid. We thank J. Clark for providing adult myocardial infarction heart samples and S. Bhattacharya for comments.

**Author Contributions** N.S. carried out the majority of experimental work and data analyses with contributions from C.A.R., A.A.D.M. and P.R.R. K.M. and R.J.S. provided the Nkx2.5Cre mouse strain and K.R.C. provided the MLC2vCre mouse strain. P.R.R. devised and planned the project and the manuscript was written by P.R.R. and N.S.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to P.R.R. (P.Riley@ich.ucl.ac.uk).



# A black hole in a globular cluster

Thomas J. Maccarone<sup>1</sup>, Arunav Kundu<sup>2</sup>, Stephen E. Zepf<sup>2</sup> & Katherine L. Rhode<sup>3,4</sup>

Globular star clusters contain thousands to millions of old stars packed within a region only tens of light years across. Their high stellar densities make it very probable that their member stars will interact or collide. There has accordingly been considerable debate about whether black holes should exist in these star clusters<sup>1–3</sup>. Some theoretical work suggests that dynamical processes in the densest inner regions of globular clusters may lead to the formation of black holes of  $\sim 1,000$  solar masses<sup>3</sup>. Other numerical simulations instead predict that stellar interactions will eject most or all of the black holes that form in globular clusters<sup>1,2</sup>. Here we report the X-ray signature of an accreting black hole in a globular cluster associated with the giant elliptical galaxy NGC 4472 (in the Virgo cluster). This object has an X-ray luminosity of about  $4 \times 10^{39} \text{ erg s}^{-1}$ , which rules out any object other than a black hole in such an old stellar population. The X-ray luminosity varies by a factor of seven in a few hours, which excludes the possibility that the object is several neutron stars superposed.

We have compared the locations of point sources found by the X-ray Multiple Mirror-Newton (XMM-Newton) satellite with the positions of globular clusters that have been identified in optical imaging data<sup>4</sup> and confirmed with optical spectroscopy (S.E.Z. *et al.*, manuscript in preparation). We have found a source at right ascension 12 h 29 min 39.7 s, declination  $+7^\circ 53' 33.4''$  (J2000), which has an average X-ray luminosity of  $2 \times 10^{39} \text{ erg s}^{-1}$  if one corrects only for foreground absorption, and  $4 \times 10^{39} \text{ erg s}^{-1}$  if one corrects also for the intrinsic absorption discussed below. Its optical counterpart (with a positional offset of about  $0.4''$ , determined by using the more accurate Chandra astrometry from ACIS-S observations on 12 June 2000) is a bright blue globular cluster, with V-band magnitude  $V = 20.99$ , and colours  $B - V = 0.68$  and  $V - R = 0.38$  (indicating a metallicity of 1/50 solar, using a colour–metallicity relation<sup>5</sup>), and has a spectroscopically determined radial velocity of  $1,477 \pm 7 \text{ km s}^{-1}$ , clearly identifying it with NGC 4472. Its optical luminosity is about  $7.5 \times 10^5$  times that of the Sun, making it one of the most luminous globular clusters in NGC 4472. This X-ray source was also detected previously by ROSAT<sup>6</sup>, but as the source lies  $6.6'$  (about 30 kpc) from the centre of NGC 4472, there have been no previous reports of searches for optical counterparts.

The globular cluster is a secure optical counterpart for the X-ray source, with a very low probability of being a chance superposition of a background active galactic nucleus (AGN) on a globular cluster. Our spectroscopically confirmed globular cluster sample includes 53 clusters in an annulus between  $6'$  and  $8'$  around the centre of NGC 4472, yielding a density of  $1.6 \times 10^{-4}$  globular clusters per square arcsecond at this distance from the centre of NGC 4472, yielding a low probability (0.005) of having a globular cluster within  $1''$  of one of the 30 brightest X-ray sources in this galaxy. In fact, only one of the 30 brightest X-ray sources is in this annulus. The space density of AGN at least as bright as the source (with an absorbed  $0.5\text{--}2.0 \text{ keV}$  flux of

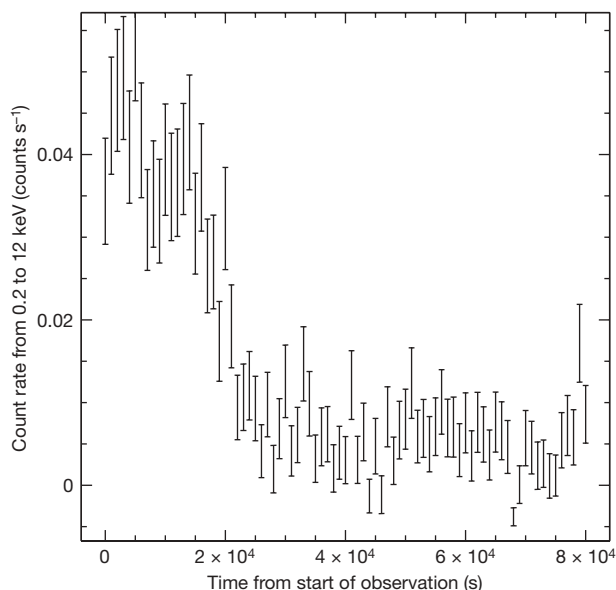
$5.6 \times 10^{-14} \text{ erg s}^{-1} \text{ cm}^{-2}$ ) is about  $5 \times 10^{-7}$  per square arcsecond (ref. 7), yielding a probability of only  $6 \times 10^{-4}$  of having such a bright AGN within  $1''$  of any of the spectroscopic globular clusters.

Furthermore, this object has a soft spectrum, and varies much more strongly at soft X-ray than at hard X-ray wavelengths. Thus, if it were a background AGN, it would have to be a narrow line Seyfert 1 galaxy, and these objects comprise only about 15% of AGN<sup>8</sup>. Finally, the AGN would have to be faint enough optically that it would neither affect the globular cluster's colours nor introduce redshifted emission lines into its spectrum. This constraint is not easily quantifiable; however, the fact that the typical broadband optical to X-ray flux ratio for AGN is larger than the ratio of fluxes for the globular cluster and the X-ray source<sup>9</sup> makes it even less likely that the X-ray source is a background AGN.

The location in a globular cluster of a given X-ray source is not surprising—about half of all X-ray sources in elliptical galaxies are in globular clusters<sup>10–12</sup>. There are, however, two remarkable features to this object—its X-ray light curve shows very high amplitude variability, and its X-ray spectrum peaks at lower energy X-rays than the spectra of typical X-ray sources in galaxies. During the first 10,000 s of the observation by the XMM-Newton satellite, the European Photon Imaging Camera (EPIC)-PN light curve shows a count rate of about  $0.04 \text{ counts s}^{-1}$  from 0.2 to 12 keV. Over the next  $\sim 10,000$  s, the count rate drops by a factor of about 7 to about  $0.006 \text{ counts s}^{-1}$ , where it remains for the next 60,000 s, when the observation ended (see Fig. 1). The same results are seen in the Metal Oxide Semiconductor cameras, but with lower count rates due to the detectors' lower effective areas. Because of this high amplitude variability, we have extracted the spectrum in two intervals, one corresponding to count rates above  $0.04 \text{ counts s}^{-1}$ , and the other to count rates below  $0.03 \text{ counts s}^{-1}$ .

In the high count rate interval, the spectrum is well fitted by a disk blackbody model, with an inner disk temperature of  $0.22 \text{ keV}$ , an inferred inner disk radius of about  $4,400 \text{ km}$ , and Galactic absorption of  $1.67 \times 10^{20} \text{ H atoms cm}^{-2}$ . The spectrum of the low count rate interval is consistent with having the same underlying continuum model, but with an increased neutral hydrogen absorbing column density ( $N_{\text{H}}$ ) of about  $3 \times 10^{21} \text{ H atoms cm}^{-2}$ ; the source varies by a factor of about 10 below  $0.7 \text{ keV}$  and is consistent with being constant above  $0.7 \text{ keV}$ . In each case, the underlying, unabsorbed luminosity is about  $4.5 \times 10^{39} \text{ erg s}^{-1}$ —the Eddington limit for accretion of hydrogen onto a 35 solar mass object, or heavier elements onto a 15 solar mass object. Similar spectral variability (that is, variations consistent with changing absorbing column density) has been reported in the past in the stellar mass black hole candidate V404 Cyg (ref. 13), although the changes in the absorption needed there were a few times larger than those reported here. The only other Galactic black hole X-ray binary that has spent a substantial amount of time above  $10^{39} \text{ erg s}^{-1}$  is GRS 1915+105, which has a foreground

<sup>1</sup>School of Physics and Astronomy, University of Southampton, Southampton SO17 1BJ, UK. <sup>2</sup>Department of Physics and Astronomy, Michigan State University, East Lansing, Michigan 48824, USA. <sup>3</sup>Department of Astronomy, Wesleyan University, Middletown, Connecticut 06459, USA. <sup>4</sup>Department of Astronomy, Yale University, New Haven, Connecticut 06520, USA.



**Figure 1 | The X-ray light curve of the globular cluster black hole candidate in NGC 4472.** The count rate from 0.2 to 12 keV is plotted versus time since the start of the observation; the error bars are  $1\sigma$  uncertainties. These data come from an XMM-Newton observation (using the EPIC-PN detector) of NGC 4472 that lasted about 100,000 s and took place on 1 January 2003. We have used the standard procedures for cleaning the XMM data to remove high background intervals (yielding about 80,000 s of good time) and to extract source spectra and light curves.

$N_{\text{H}}$  of about  $5 \times 10^{22}$  H atoms  $\text{cm}^{-2}$  (ref. 14), so we would not expect to be able to see  $N_{\text{H}}$  variations in GRS 1915+105 that are as small as the ones we are reporting in the NGC 4472 source.

This large-amplitude variability is obviously inconsistent with the idea that the X-ray emission in this source could come from a superposition of bright neutron stars—and in fact, this variability is one of the only ways to prove that an extragalactic globular cluster X-ray source is a black hole<sup>15</sup>. If the inner disk radius of the accretion disk around the X-ray source is 4,400 km, and this corresponds to the innermost stable circular orbit of a Schwarzschild black hole, then the black hole mass is at least about 400 solar masses, so it is plausible that this object is an accreting intermediate mass black hole. The key remaining issue is whether the best fitting inner disk radius really is the innermost stable circular orbit around the black hole.

Proposals exist for mechanisms capable of producing soft X-ray spectra at high luminosities from stellar-mass, rather than intermediate-mass, black holes. At very high accretion rates, slim disks (that is, disks with scale heights of about 10% of their radii) should form<sup>16</sup> around black holes, and these should have very large, optically thick photospheres. The source spectrum that we see is consistent with the emission expected from a 10 solar mass black hole accreting at a rate of about  $3 \times 10^{-7}$  solar masses per year in the context of such models<sup>17,18</sup>. The variability in the source discussed here is predominantly in soft X-rays, in direct contrast to what was seen in the M101 source first suggested as an example of this mode of accretion<sup>17</sup>—in such outflow models, the variability should be due to changes in the size of the photosphere, not in the absorption column in front of it. On the other hand, as V404 Cyg did show a similar X-ray spectrum<sup>19</sup> and similar spectral variability<sup>13</sup>, and at a luminosity only slightly lower, it seems quite plausible that the object in NGC 4472 is a super-Eddington stellar mass black hole; perhaps the M101 source is at a lower inclination angle than V404 Cyg and the NGC 4472 black hole, so it is less obscured by the outflow generated through the super-Eddington accretion. Thus the case for an intermediate mass black hole in this cluster is inconclusive, in contrast to the strong case for a black hole of some mass rather than a neutron star.

It has also been suggested that it may be possible to exceed the Eddington limit by a large factor in slim disk models<sup>18</sup>. The luminosity may be Eddington-limited at every radius of an accretion disk. The actual brightness will then be the Eddington limit times the logarithm of the ratio of the mass accretion rate to the Eddington mass accretion rate under the assumption of standard radiatively efficient accretion. The unabsorbed luminosity is then 20–26 times the Eddington rate for hydrogen accretion onto a 1.4 solar mass neutron star. This yields an outer disk mass accretion rate of at least 5 solar masses per year. Additionally, slim disks cannot exist around neutron stars, at least in the present models, as it will not be possible to advect trapped photons in the case of a neutron star accretor. Thus, even in the high accretion rate slim disk case, the identification of this source as one with a black hole accretor is secure. The other existing mechanism for making neutron stars accrete far above their Eddington limit is polar accretion in X-ray pulsars, such as SMC X-1, which exceeds its Eddington limit by a factor of about 15 (ref. 20): but this works only in young stellar populations, because magnetic fields in accreting objects decay over time. Therefore, it is highly unlikely an accretion-powered pulsar will be a globular cluster source.

The present discovery allows us to place some limited constraints on models of the formation and evolution of systems with accreting black holes in globular clusters. As this object has been bright at least since the ROSAT observations in 1994 (ref. 6), and was detected in a Chandra observation and an XMM observation separated by about 3 years, it seems clear that the outburst duration is at least a decade (or that the source is persistent). In the context of models with stellar mass black hole accretors in globular clusters, this object is most probably in a system with a red giant donor star in a long (that is at least one-month) orbital period<sup>21,22</sup>. It has been suggested that there should typically be a single stellar mass black hole in most globular clusters, and that most of these black holes should be in wide binaries, leading to low X-ray bright duty cycles<sup>15</sup>. At present, with only a single confirmed black hole in the  $\sim 6,000$  globular clusters of NGC 4472 (ref. 4), we can estimate that the duty cycle of bright X-ray emission must not be much less than 1/6,000, or 1/400 if one includes only the spectroscopically selected clusters. In either case, this is consistent with previous suggestions that the duty cycle as bright X-ray sources of these systems should be no more than about 1/1,000. Attempts to understand the population synthesis and stellar evolution of accreting binaries with intermediate mass black holes have also been made<sup>23–26</sup>. Theoretical predictions of the probability that the black hole will capture a companion, and of the probability that a cluster will contain an intermediate mass black hole, are highly uncertain. Therefore, it is difficult to place constraints on theory with only a single object, which may or may not be of intermediate mass.

Received 18 September; accepted 7 November 2006.

Published online 3 January 2007.

1. Kulkarni, S. R., Hut, P. & McMillan, S. Stellar black holes in globular clusters. *Nature* **364**, 421–423 (1993).
2. Sigurdsson, S. & Hernquist, L. Primordial black holes in globular clusters. *Nature* **364**, 423–425 (1993).
3. Miller, M. C. & Hamilton, D. P. Production of intermediate-mass black holes in globular clusters. *Mon. Not. R. Astron. Soc.* **330**, 232–240 (2002).
4. Rhode, K. L. & Zepf, S. E. The globular cluster system in the outer regions of NGC 4472. *Astron. J.* **121**, 210–224 (2001).
5. Smits, M., Maccarone, T. J., Kundu, A. & Zepf, S. E. The globular cluster mass/low mass X-ray binary correlation: implications for kick velocity distributions from supernovae. *Astron. Astrophys.* **458**, 477–484 (2006).
6. Colbert, E. J. M. & Ptak, A. F. A catalog of candidate intermediate-luminosity X-ray objects. *Astrophys. J. Suppl. Ser.* **143**, 25–45 (2002).
7. Mushotzky, R. F., Cowie, L. L., Barger, A. J. & Arnaud, K. A. Resolving the extragalactic hard X-ray background. *Nature* **404**, 459–464 (2000).
8. Williams, R. J., Pogge, R. W. & Mathur, S. Narrow-line Seyfert 1 galaxies from the Sloan Digital Sky Survey early data release. *Astron. J.* **124**, 3042–3049 (2002).
9. Maccacaro, T., Gioia, I. M., Wolter, A., Zamorani, G. & Stocke, J. T. The X-ray spectra of the extragalactic sources in the Einstein extended medium sensitivity survey. *Astrophys. J.* **326**, 680–690 (1988).
10. Angelini, L., Loewenstein, M. & Mushotzky, R. F. The X-ray globular cluster population in NGC 1399. *Astrophys. J.* **557**, L35–L38 (2001).



11. Kundu, A., Maccarone, T. J. & Zepf, S. E. The low-mass X-ray binary-globular cluster connection in NGC 4472. *Astrophys. J.* **574**, L5–L9 (2002).
12. Maccarone, T. J., Kundu, A. & Zepf, S. E. The low-mass X-ray binary-globular cluster connection. II. NGC 4472 X-ray source properties and source catalogs. *Astrophys. J.* **586**, 814–825 (2003).
13. Oosterbroek, T. *et al.* Spectral and timing behaviour of GS 2023+338. *Astron. Astrophys.* **321**, 776–790 (1997).
14. Lee, J. C. *et al.* High-resolution Chandra HETGS and Rossi X-Ray Timing Explorer observations of GRS 1915+105: A hot disk atmosphere and cold gas enriched in iron and silicon. *Astrophys. J.* **567**, 1102–1111 (2002).
15. Kalogera, V., King, A. R. & Rasio, F. A. Could black hole X-ray binaries be detected in globular clusters? *Astrophys. J.* **601**, L171–L174 (2004).
16. Abramowicz, M. A., Czerny, B., Lasota, J. P. & Szuszkiewicz, E. Slim accretion discs. *Astrophys. J.* **332**, 646–658 (1988).
17. Mukai, K., Pence, W. D., Snowden, S. L. & Kuntz, K. D. Chandra observation of luminous and ultraluminous X-ray binaries in M101. *Astrophys. J.* **582**, 184–189 (2003).
18. King, A. R. & Pounds, K. A. Black hole winds. *Mon. Not. R. Astron. Soc.* **345**, 657–659 (2003).
19. Zycki, P., Done, C. & Smith, D. A. The 1989 May outburst of the soft X-ray transient GS 2023+338 (V404 Cyg). *Mon. Not. R. Astron. Soc.* **309**, 561–575 (1999).
20. Liu, Q. Z., van Paradijs, J. & van den Heuvel, E. P. J. A catalogue of high mass X-ray binaries. *Astron. Astrophys. Suppl.* **147**, 25–49 (2000).
21. Truss, M. & Done, C. The decline and fall of GRS 1915+105: the end is nigh? *Mon. Not. R. Astron. Soc.* **368**, L25–L29 (2006).
22. Portegies Zwart, S. F., Dewi, J. & Maccarone, T. Intermediate mass black holes in accreting binaries: formation, evolution and observational appearance. *Mon. Not. R. Astron. Soc.* **355**, 413–423 (2004).
23. Blecha, L. *et al.* Close binary interactions of intermediate-mass black holes: Possible ultraluminous X-ray sources? *Astrophys. J.* **642**, 427–437 (2006).
24. Hopman, C., Portegies Zwart, S. & Alexander, T. Ultraluminous X-ray sources as intermediate-mass black holes fed by tidally captured stars. *Astrophys. J.* **604**, L101–L104 (2004).
25. Madhusudhan, N. *et al.* Models of ultraluminous X-ray sources with intermediate-mass black holes. *Astrophys. J.* **640**, 918–922 (2006).
26. Patruno, A., Colpi, M., Faulkner, A. & Possenti, A. Radio pulsars around intermediate-mass black holes in superstellar clusters. *Mon. Not. R. Astron. Soc.* **364**, 344–352 (2005).

**Acknowledgements** K.L.R. is an NSF Astronomy and Astrophysics Postdoctoral Fellow. We thank E. Koerding, T. Dwelly, S. Jester, J. Salzer and G. Bergond for useful communications.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to T.J.M. (tjm@phys.soton.ac.uk).

## LETTERS

# Spin correlations in the electron-doped high-transition-temperature superconductor $\text{Nd}_{2-x}\text{Ce}_x\text{CuO}_{4\pm\delta}$

E. M. Motoyama<sup>1</sup>, G. Yu<sup>1</sup>, I. M. Vishik<sup>1</sup>, O. P. Vajk<sup>2</sup>, P. K. Mang<sup>3</sup> & M. Greven<sup>3,4</sup>

High-transition-temperature (high- $T_c$ ) superconductivity develops near antiferromagnetic phases, and it is possible that magnetic excitations contribute to the superconducting pairing mechanism. To assess the role of antiferromagnetism, it is essential to understand the doping and temperature dependence of the two-dimensional antiferromagnetic spin correlations. The phase diagram is asymmetric with respect to electron and hole doping, and for the comparatively less-studied electron-doped materials, the antiferromagnetic phase extends much further with doping<sup>1,2</sup> and appears to overlap with the superconducting phase. The archetypal electron-doped compound  $\text{Nd}_{2-x}\text{Ce}_x\text{CuO}_{4\pm\delta}$  (NCCO) shows bulk superconductivity above  $x \approx 0.13$  (refs 3, 4), while evidence for antiferromagnetic order has been found up to  $x \approx 0.17$  (refs 2, 5, 6). Here we report inelastic magnetic neutron-scattering measurements that point to the distinct possibility that genuine long-range antiferromagnetism and superconductivity do not coexist. The data reveal a magnetic quantum critical point where superconductivity first appears, consistent with an exotic quantum phase transition between the two phases<sup>7</sup>. We also demonstrate that the pseudogap phenomenon in the electron-doped materials, which is associated with pronounced charge anomalies<sup>8–11</sup>, arises from a build-up of spin correlations, in agreement with recent theoretical proposals<sup>12,13</sup>.

In their as-grown state, the electron-doped materials exhibit antiferromagnetic (AF) order throughout the accessible doping range, and an oxygen reduction treatment is required to induce superconductivity<sup>3</sup>. Previous inelastic neutron scattering experiments on as-grown, non-superconducting (non-SC) NCCO demonstrated<sup>6</sup> that the two-dimensional (2D) spin-correlation length  $\xi$  observed above the Néel temperature is exponentially dependent on inverse temperature:

$$\xi(x, T) = A(x) \exp(2\pi\rho_s(x)/T) \quad (1)$$

for cerium concentrations ranging from zero up to the solubility limit of  $x \approx 0.18$ . This behaviour indicates the existence of an underlying ground state with long-range 2D AF order. Owing to weak spin-space anisotropies and three-dimensional couplings, NCCO exhibits three-dimensional AF order at a non-zero Néel temperature, as observed in the elastic scattering channel<sup>2,6</sup>. The spin stiffness  $\rho_s(x)$  decreases monotonically with increasing electron concentration, with  $\rho_s(0.18)/\rho_s(0) \approx 25\%$  (ref. 6), and the doping dependence of  $\rho_s(x)$  and of the amplitude  $A(x)$  is remarkably close to that for the randomly-diluted spin-one-half square-lattice Heisenberg antiferromagnet  $\text{La}_2\text{Cu}_{1-z}(\text{Zn}, \text{Mg})_z\text{O}_4$  (ref. 14).

Magnetic inelastic neutron-scattering experiments in the SC phase have become possible only in recent years<sup>15,16</sup>. We have carried out

two-axis measurements of the spin correlations in eight oxygen-reduced NCCO crystals in the cerium concentration range  $0.038 \leq x \leq 0.154$  (Fig. 1). The data are fitted to a 2D lorentzian,  $S(q_{2D}) = S(0)/(1 + q_{2D}^2\xi^2)$ , convoluted with the calculated instrumental resolution, where  $q_{2D}$  is the distance in momentum space from the 2D AF zone centre (Fig. 2). The non-SC samples ( $x \leq 0.129$ ) follow equation (1), consistent with bulk AF order in the ground state (Fig. 3). Although this behaviour is qualitatively the same as that for as-grown NCCO<sup>6</sup>, the spin stiffness decreases much more rapidly with doping. The data for the SC sample with  $x = 0.134$  are fitted to equation (1) with a small value of the spin stiffness  $\rho_s$  but are equally well described by the simple power law  $\xi \propto 1/T^{\nu_T}$  with exponent  $\nu_T = 1.0(5)$ . The power-law behaviour, indicated by the dashed curve in Fig. 3, would imply that  $\rho_s$  is already zero and that the system is quantum critical at this cerium concentration. Figure 4a demonstrates that  $\rho_s$  approaches zero at  $x_{\text{AF}} = 0.134(4)$  in an approximately linear fashion. In a fundamental departure from the above behaviour, we find that in the SC samples with  $x \geq 0.145$ ,  $\xi$  remains finite down to the lowest temperatures. The low-temperature correlation length  $\xi_0$  for these samples increases as  $x_{\text{AF}}$  is approached from above (Fig. 4a).

Previous elastic neutron measurements<sup>2,4,6</sup> indicated that oxygen-reduced NCCO exhibits AF order up to  $x \approx 0.17$ . Such measurements in our crystals indeed reveal Néel order. However, our inelastic results demonstrate that, contrary to previous belief<sup>6,11,17</sup>, the ground state of SC samples exhibits only short-range spin correlations. Moreover, prior inelastic neutron-scattering experiments clearly revealed a SC magnetic gap, despite the presence of AF Bragg peaks in the elastic response<sup>15,16</sup>. We conclude that the AF phase boundary in fact terminates at  $x_{\text{AF}} = 0.134(4)$ , and that magnetic Bragg peaks observed at higher cerium concentrations originate from regions of the samples that were not fully oxygen-annealed. While a relatively small volume fraction of such macroscopic remnants of the AF as-grown state can give rise to significant Bragg scattering, our inelastic measurements are fortunately insensitive to their presence. This conclusion is consistent with the observation that the Néel transition is very broad in SC samples (Fig. 4b), and also with muon spin-resonance results<sup>4</sup>, which show a significant decrease of the AF volume fraction near  $x = 0.14$ . We note that the spurious elastic signal from the remnant AF regions should not be confused with the spurious elastic signal in a magnetic field<sup>18</sup> due to the paramagnetic decomposition product  $(\text{Nd}, \text{Ce})_2\text{O}_3$ .

For as-grown NCCO<sup>6</sup> and for  $\text{La}_2\text{Cu}_{1-z}(\text{Zn}, \text{Mg})_z\text{O}_4$  (ref. 14) it was found that, to a very good approximation, the Néel temperature  $T_N(x)$  is a contour of constant 2D correlation length with

<sup>1</sup>Department of Physics, Stanford University, Stanford, California 94305, USA. <sup>2</sup>NIST Center for Neutron Research, National Institute of Standards and Technology, Gaithersburg, Maryland 20899, USA. <sup>3</sup>Department of Applied Physics, Stanford University, Stanford, California 94305, USA. <sup>4</sup>Stanford Synchrotron Radiation Laboratory, Stanford, California 94309, USA.

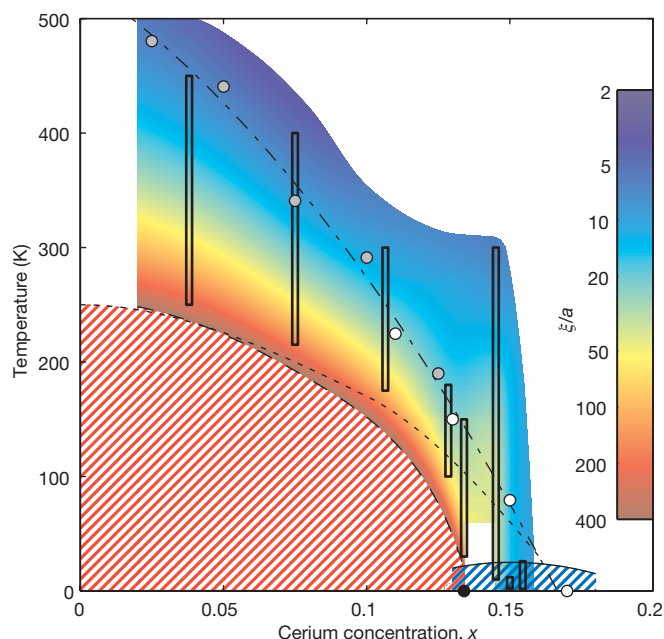


$\xi/a = 200$ –400 and 100, respectively. Following the observations for as-grown NCCO, we plot the extrapolated contour of  $\xi/a = 400$  as a dashed curve in Fig. 1. This estimate of the underlying bulk Néel temperature coincides with the measured  $T_N$  at  $x = 0.038$ , but it lies systematically lower at higher cerium concentrations, approaching  $T_N = 0$  at  $x_{AF} \approx 0.134$ .

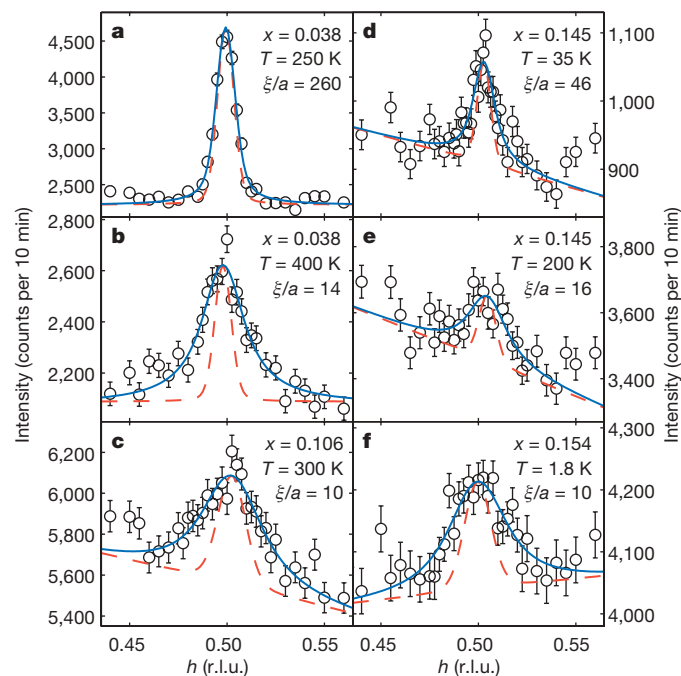
The decrease to zero of the spin stiffness at  $x_{AF} \approx 0.134$  and the finite values of  $\xi_0$  for  $x > x_{AF}$  indicates a fundamental change in the nature of the magnetic ground state. The contribution of the AF remnants may lead to a slight over-estimate of the spin correlations, and consequently of  $\rho_s$  and  $\xi_0$ , but we emphasize that the qualitative change in behaviour is a robust result. The NCCO phase diagram resembles those of other unconventional superconductors, such as the heavy-fermion compound CeRhIn<sub>5</sub>, in which the AF and SC phases are believed to be separated by a first-order boundary<sup>19</sup>. Although we cannot rule out a genuine underlying coexistence between AF and SC order, such coexistence would be confined to rather a narrow doping range. However, the behaviour of  $\rho_s(x)$ , which decreases continuously by more than an order of magnitude with doping, together with the crossover to power-law behaviour of  $\xi(x, T)$  near  $x = 0.134$ , suggests another scenario: a second-order quantum phase transition between the AF and SC phases. This quantum phase transition would be described by a dynamic critical

exponent of  $z = 1/\nu_T \approx 1.0(5)$ , which differs from the value  $z = 2$  predicted for a transition from the antiferromagnet to a non-SC paramagnet<sup>12</sup>. If hyperscaling holds, the spin stiffness for a 2D system is expected to decrease as  $\rho_s \propto (x_{AF} - x)^{\nu_0 z}$  where  $\nu_0$  is the exponent describing the divergence of  $\xi_0$  as  $x_{AF}$  is approached from above. From the approximately linear behaviour of the spin stiffness, we therefore have  $\nu_0 \approx 1$ . We cannot independently determine  $\nu_0$ , because we do not have sufficient information for  $\xi_0(x)$ . It is also possible that the system lies above the upper critical dimension, in which case mean-field behaviour with  $\rho_s \propto (x_{AF} - x)^{2\beta_{\text{mean field}}}$  is expected.  $\beta_{\text{mean field}} = 1/2$ , so this is consistent with the observed behaviour.

Our results for  $\xi(x, T)$  also have important consequences for the relationship between AF correlations and the pseudogap physics in the electron-doped copper oxides, which appears to be different from



**Figure 1 | The temperature–doping phase diagram for oxygen-reduced NCCO.** The red and blue hashed areas indicate long-range AF order and superconductivity, respectively. The black circle at zero temperature indicates the approximate location of a magnetic quantum phase transition. The instantaneous 2D spin-correlation length  $\xi(x, T)$  in the CuO<sub>2</sub> sheets was measured at the doping levels and over the temperature ranges indicated by the vertical bars. The colour scale shows  $\xi$ , in units of the planar lattice constant  $a$ , interpolated and extrapolated from the measured values. The Néel temperature  $T_N$  is shown as the dotted curve, while the dashed curve is the extrapolated contour of  $\xi/a = 400$ . The measurement of  $T_N$  is contaminated by remnants of the as-grown state of NCCO, so that the true AF phase extends only to  $x_{AF} \approx 0.13$ , close to where superconductivity first appears. This is established from the fact that  $\xi$  diverges exponentially upon cooling for non-superconducting compositions at lower electron concentrations, while it remains finite in superconducting samples. The small remaining overlap indicated in the figure may be caused by cerium and oxygen inhomogeneities. The grey and white circles indicate optical conductivity measurements of the pseudogap temperature  $T^*$  on NCCO crystals<sup>9</sup> and Pr<sub>2-x</sub>Ce<sub>x</sub>CuO<sub>4±δ</sub> thin films<sup>11</sup>, respectively. The dot-dashed curve is a guide to the eye.

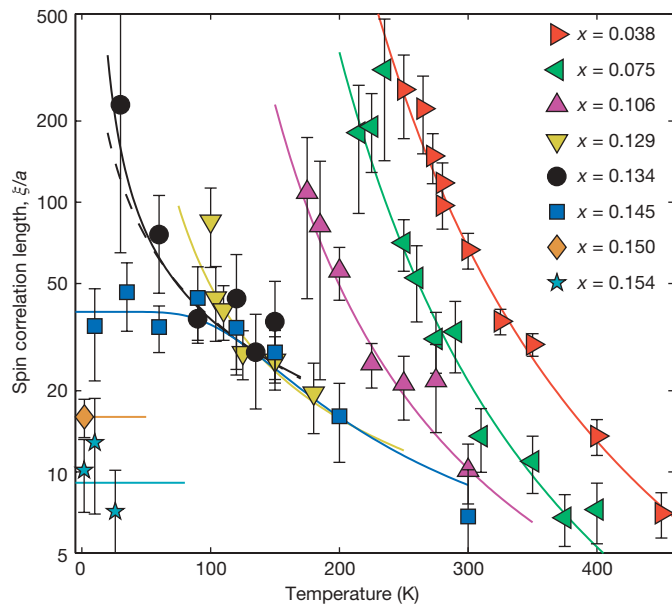


**Figure 2 | Representative two-axis scans used to measure the spin correlation length.** The scans are along  $(h, h)$  about the 2D AF zone centre  $(1/2, 1/2)$  and are fitted (solid blue curves) to a 2D Lorentzian convoluted with the calculated instrumental resolution (dashed red curves). Shown are data at **a**,  $T = 250$  K  $\approx T_N$  and **b**,  $T = 400$  K for  $x = 0.038$ ; **c**,  $T = 300$  K for  $x = 0.106$ ; **d**,  $T = 35$  K and **e**,  $T = 200$  K for  $x = 0.145$ ; and **f**,  $T = 1.8$  K for  $x = 0.154$ . Wave vectors are represented as  $(h, k, l)$  in reciprocal lattice units (r.l.u.), where  $Q = (2\pi h/a, 2\pi k/a, 2\pi l/c)$  is the momentum transfer, and  $a$  and  $c$  are the lattice constants of the tetragonal system (space group  $I4/mmm$ ; for  $x = 0.038$ , for example, the room-temperature lattice constants are  $a = 3.93$  Å and  $c = 12.09$  Å). Vertical error bars represent uncertainties ( $1\sigma$ ) assuming Poisson statistics. The measurements were performed in two-axis mode on the BT2 and BT9 triple-axis spectrometers at the NIST Center for Neutron Research. The incident neutron energy was  $E_i = 14.7$  meV. In previous experiments on La<sub>2</sub>Cu<sub>1-x</sub>(Zn,Mg)<sub>2</sub>O<sub>4</sub><sup>14</sup> and as-grown NCCO<sup>9</sup>, this energy proved to be sufficiently large in the temperature region  $T_N < T < 2T_N$  to reliably extract the instantaneous structure factor  $S(Q)$ . The collimations were: **a**, **b**,  $40^\circ$ – $23^\circ$ –sample– $20^\circ$ ; **c**,  $60^\circ$ – $40^\circ$ –sample– $40^\circ$ ; **d**,  $40^\circ$ – $47^\circ$ –sample– $10^\circ$ ; **e**,  $40^\circ$ – $47^\circ$ –sample– $20^\circ$ ; and **f**,  $40^\circ$ – $47^\circ$ –sample– $40^\circ$ . The NCCO crystals were grown in 4 atm of oxygen using the travelling-solvent floating-zone technique, and subsequently annealed for 10 h at  $970^\circ$  C in argon, followed by 20 h at  $500^\circ$  C in oxygen. The sample masses range from 1 to 5 g. The oxygen reduction treatment, required for superconductivity to appear, is a non-equilibrium process resulting in unavoidable oxygen inhomogeneities. Cerium concentrations  $x$  were determined from inductively coupled plasma (ICP) spectroscopy, with typical variation of  $\Delta x \approx 0.005$  along the growth direction. Superconductivity is observed from magnetic susceptibility measurements for  $x \geq 0.134$ .

that of the hole-doped materials<sup>9,12,13</sup>. The pseudogap (charge anomalies associated with the opening of a partial gap along the Fermi surface) has been discerned in photoemission<sup>8,10</sup>, optical spectroscopy<sup>9,11</sup>, and charge transport<sup>9</sup> experiments on NCCO crystals and  $\text{Pr}_{2-x}\text{Ce}_x\text{CuO}_{4\pm\delta}$  films up to  $x = 0.15$ <sup>8,11</sup>. First, our finding that the AF phase terminates at  $x_{\text{AF}} = 0.134(4)$  refutes statements that a possible quantum phase transition at  $x \approx 0.17$  is related to the disappearance of AF order<sup>11,17</sup>. Second, we find that the spin-correlation length changes smoothly across the pseudogap temperature  $T^*$  and, remarkably, up to  $x = 0.145$  along  $T^*(x)$  it follows the simple relationship:

$$\frac{\xi^*}{a} = \frac{C}{x_c - x} \quad (2)$$

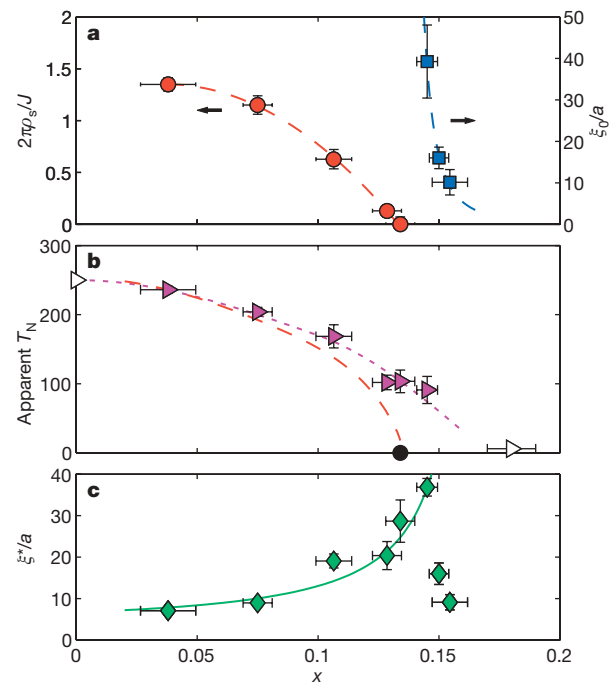
with fitting parameters  $C = 0.96(12)$  and  $x_c = 0.171(4)$ . One interpretation of the pseudogap is that it signifies a change in the spin scattering of the electrons as the AF correlations exceed the carrier mean free path<sup>9</sup> ( $l$ ) or the single-particle thermal de Broglie wavelength<sup>12</sup> ( $\xi_{\text{th}} = \hbar v_F / \pi k_B T$ ) upon cooling. Boltzmann transport theory, which might be expected to hold qualitatively if  $l > a$  and  $T > T^*$ , applied to direct-current resistivity data for NCCO yields  $l(T^*) \approx 2.4a$  and  $l(T^*) \approx 25a$  for  $x = 0.125$  and  $x = 0.15$ , respectively<sup>9</sup>. This trend is qualitatively consistent with our results up to  $x = 0.145$  seen in Fig. 4c. However, given the approximately linear relationship  $T^* \propto (x^* - x)$  (Fig. 1), the thermal de Broglie wavelength at  $T^*$  exhibits the same quantitative doping dependence as does equation (2). Using the value  $v_F = 2.2 \times 10^7 \text{ cm s}^{-1}$  for the bare Fermi velocity<sup>12</sup>, we find that  $\xi^* = 2.6(2)\xi_{\text{th}}$ . We conclude that  $T^*$  is a crossover temperature below which the spin correlations become longer than the thermal de Broglie wavelength, in agreement with theoretical work<sup>12</sup>. The pseudogap phenomenon in the electron-doped copper oxides therefore results from 2D AF spin correlations, and does not appear to be a direct precursor to superconductivity. That  $\xi^* \propto \xi_{\text{th}}$  is obeyed so well suggests that the contributions from the remnant AF regions to our measured  $\xi(x, T)$  are negligible.



**Figure 3 | The temperature dependence of the spin correlation length at various cerium concentrations.** Vertical bars represent uncertainties of  $1\sigma$ . The data for  $x \leq 0.134$  are fitted to equation (1). The spin stiffness may already be zero for  $x = 0.134$ , because a fit to a simple power law  $\xi \propto 1/T$  describes the data equally well (dashed curve); power-law behaviour is expected at a quantum critical point. For  $x = 0.145$  and higher,  $\xi$  does not diverge, but instead remains finite at low temperatures, demonstrating the absence of genuine long-range AF order. The curves drawn for these latter data are guides to the eye. Superconductivity is observed for  $x \geq 0.134$ .

Equation (2) would suggest that  $\xi^*$  diverges at  $x_c \approx x^*$ . However, as seen from Fig. 4c, this relation breaks down near optimal doping, presumably owing to the emergence of a new length scale that limits the development of spin correlations. Interestingly, at optimal doping ( $x = 0.15$ ),  $\xi^* \approx \xi_0$  is comparable to the SC coherence length<sup>20</sup>  $\xi_{\text{SC}} = 58 \text{ \AA} \approx 15a$ . The spin correlations near optimal doping are still relatively large, consistent with suggestions based on Raman scattering<sup>21</sup> and photoemission<sup>22</sup> that the  $d_{x^2-y^2}$  SC order parameter is non-monotonic due to AF fluctuations.

The new experimental results for the magnetic phase diagram and pseudogap physics contain important implications for theories of high- $T_c$  superconductivity. By avoiding spurious scattering that significantly contaminates the elastic response, we have established from measurements of the 2D spin correlations that genuine coexistence of AF and SC order is essentially absent in  $\text{Nd}_{2-x}\text{Ce}_x\text{CuO}_{4\pm\delta}$ . On symmetry grounds, a possible second-order quantum phase transition between these two types of order would seem unlikely and exotic. One scenario is that an underlying first-order transition is rendered second-order owing to microscopic disorder<sup>23</sup>. Such



**Figure 4 | Spin stiffness, spin correlations at low-temperature, apparent Néel temperature, and spin correlations along  $T^*$ .** **a**, Doping dependence of the spin stiffness  $\rho_s$  (plotted as  $2\pi\rho_s/J$ , where  $J = 125 \text{ meV}$  is the AF superexchange for the undoped Mott insulator  $\text{Nd}_2\text{CuO}_4$ <sup>2,6</sup>) and of the low-temperature spin correlation length  $\xi_0$ . Vertical error bars here represent uncertainties of  $1\sigma$ . Horizontal error bars in all panels represent the measured range of cerium concentration in each crystal. Dashed curves are guides to the eye. The spin stiffness decreases smoothly with doping and reaches zero in an approximately linear fashion around  $x_{\text{AF}} \approx 0.134$ . The ground state for  $x < x_{\text{AF}}$  has long-range AF order, whereas long-range order is absent for  $x > x_{\text{AF}}$ , as seen from the finite values of  $\xi_0$ . The doping dependence of  $\xi_0$  indicates a divergence as the critical point is approached from the right. **b**, Apparent Néel temperature  $T_N$ , as determined from elastic scattering, as a function of doping. The temperature dependence of the measured order parameter (not shown) was modelled using a gaussian distribution of  $T_N$ , and the vertical bars indicate the full-width at half-maximum (FWHM) of this distribution. Measurements were not performed on all samples; previous data<sup>6</sup> are indicated by open symbols. The dotted and dashed curves are the same as in Fig. 1. **c**, The spin correlation length  $\xi^*$  measured at or extrapolated to the pseudogap temperature  $T^*$ . Vertical bars represent uncertainties of  $1\sigma$ . Below optimal doping ( $x < 0.15$ ),  $\xi^*$  is given by the single-particle thermal de Broglie wavelength and increases as  $\xi^* \propto 1/(x^* - x)$  (fitted curve). However, this relationship breaks down near optimal doping, where  $\xi^*$  is found not to exceed the SC coherence length.



disorder is found in most high- $T_c$  superconductors<sup>24</sup> and, in the present case, it might be the randomness associated with the Nd–Ce substitution. Alternatively, such a phase transition may be an example of ‘deconfined’ quantum criticality<sup>7</sup>, a new paradigm for quantum phase transitions. To further elucidate the nature of the transition from AF to SC order, a detailed complementary study of the superconducting critical properties on small samples with minimal oxygen and cerium inhomogeneities would be desirable.

Received 28 June; accepted 9 November 2006.

- Keimer, B. *et al.* Magnetic excitations in pure, lightly doped, and weakly metallic  $\text{La}_2\text{CuO}_4$ . *Phys. Rev. B* **46**, 14034–14053 (1992).
- Matsuda, M. *et al.* Magnetic order, spin correlations, and superconductivity in single-crystal  $\text{Nd}_{2-x}\text{Ce}_x\text{CuO}_{4+\delta}$ . *Phys. Rev. B* **45**, 12548–12554 (1992).
- Takagi, H., Uchida, S. & Tokura, Y. Superconductivity produced by electron doping in  $\text{CuO}_2$ -layered compounds. *Phys. Rev. Lett.* **62**, 1197–1200 (1989).
- Uefuji, T. *et al.* Coexistence of antiferromagnetic ordering and high- $T_c$  superconductivity in electron-doped superconductor  $\text{Nd}_{2-x}\text{Ce}_x\text{CuO}_4$ . *Physica C* **357–360**, 208–211 (2001).
- Uefuji, T., Kurahashi, K., Fujita, M., Matsuda, M. & Yamada, K. Electron-doping effect on magnetic order and superconductivity in  $\text{Nd}_{2-x}\text{Ce}_x\text{CuO}_4$  single crystals. *Physica C* **378–381**, 273–277 (2002).
- Mang, P. K., Vajk, O. P., Arvanitaki, A., Lynn, J. W. & Greven, M. Spin correlations and magnetic order in nonsuperconducting  $\text{Nd}_{2-x}\text{Ce}_x\text{CuO}_{4\pm\delta}$ . *Phys. Rev. Lett.* **93**, 027002 (2004).
- Senthil, T., Vishwanath, A., Balents, L., Sachdev, S. & Fisher, M. P. A. Deconfined quantum critical points. *Science* **303**, 1490–1494 (2004).
- Armitage, N. P. *et al.* Doping dependence of an  $n$ -type cuprate superconductor investigated by angle-resolved photoemission spectroscopy. *Phys. Rev. Lett.* **88**, 257001 (2002).
- Onose, Y., Taguchi, Y., Ishizaka, K. & Tokura, Y. Charge dynamics in underdoped  $\text{Nd}_{2-x}\text{Ce}_x\text{CuO}_4$ : Pseudogap and related phenomena. *Phys. Rev. B* **69**, 024504 (2004).
- Matsui, H. *et al.* Angle-resolved photoemission spectroscopy of the antiferromagnetic superconductor  $\text{Nd}_{1.87}\text{Ce}_{0.13}\text{CuO}_4$ : Anisotropic spin-correlation gap, pseudogap, and the induced quasiparticle mass enhancement. *Phys. Rev. Lett.* **94**, 047005 (2005).
- Zimmers, A. *et al.* Infrared properties of electron-doped cuprates: Tracking normal-state gaps and quantum critical behavior in  $\text{Pr}_{2-x}\text{Ce}_x\text{CuO}_4$ . *Europhys. Lett.* **70**, 225–231 (2005).
- Kyung, B., Hankevych, V., Daré, A.-M. & Tremblay, A.-M. S. Pseudogap and spin fluctuations in the normal state of the electron-doped cuprates. *Phys. Rev. Lett.* **93**, 147004 (2004).
- Markiewicz, R. S. Mode-coupling model of Mott gap collapse in the cuprates: Natural phase boundary for quantum critical points. *Phys. Rev. B* **70**, 174518 (2004).
- Vajk, O. P., Mang, P. K., Greven, M., Gehring, P. M. & Lynn, J. W. Quantum impurities in the two-dimensional spin one-half Heisenberg antiferromagnet. *Science* **295**, 1691–1695 (2002).
- Yamada, K. *et al.* Commensurate spin dynamics in the superconducting state of an electron-doped cuprate superconductor. *Phys. Rev. Lett.* **90**, 137004 (2003).
- Motoyama, E. M. *et al.* Magnetic field effect on the superconducting magnetic gap of  $\text{Nd}_{1.85}\text{Ce}_{0.15}\text{CuO}_4$ . *Phys. Rev. Lett.* **96**, 137002 (2006).
- Dagan, Y. *et al.* Origin of the anomalous low temperature upturn in the resistivity of the electron-doped cuprate superconductors. *Phys. Rev. Lett.* **94**, 057005 (2005).
- Mang, P. K. *et al.* Phase decomposition and chemical inhomogeneity in  $\text{Nd}_{2-x}\text{Ce}_x\text{CuO}_{4\pm\delta}$ . *Phys. Rev. B* **70**, 094507 (2004).
- Park, T. *et al.* Hidden magnetism and quantum criticality in the heavy fermion superconductor  $\text{CeRhIn}_5$ . *Nature* **440**, 65–68 (2006).
- Wang, Y. *et al.* Dependence of upper critical field and pairing strength on doping in cuprates. *Science* **299**, 86–89 (2003).
- Blumberg, G. *et al.* Nonmonotonic  $d_{x^2-y^2}$  superconducting order parameter in  $\text{Nd}_{2-x}\text{Ce}_x\text{CuO}_4$ . *Phys. Rev. Lett.* **88**, 107002 (2002).
- Matsui, H. *et al.* Direct observation of a nonmonotonic  $d_{x^2-y^2}$  -wave superconducting gap in the electron-doped high- $T_c$  superconductor  $\text{Pr}_{0.89}\text{LaCe}_{0.11}\text{CuO}_4$ . *Phys. Rev. Lett.* **95**, 017003 (2005).
- Aizenman, M. & Wehr, J. Rounding effects of quenched randomness on first-order phase transitions. *Commun. Math. Phys.* **130**, 489–528 (1990).
- Eisaki, H. *et al.* Effect of chemical inhomogeneity in bismuth-based copper oxide superconductors. *Phys. Rev. B* **69**, 064512 (2004).

**Acknowledgements** We thank N. Bontemps, S. Chakravarty, S. A. Kivelson, R. S. Markiewicz and A.-M. S. Tremblay for discussions. The work at Stanford University was supported by grants from the Department of Energy and the National Science Foundation. E.M.M. acknowledges support through the NSF Graduate Fellowship programme.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.G. ([greven@stanford.edu](mailto:greven@stanford.edu)).

## LETTERS

# Quantification of actinide $\alpha$ -radiation damage in minerals and ceramics

Ian Farnan<sup>1</sup>, Herman Cho<sup>2</sup> & William J. Weber<sup>2</sup>

There are large amounts of heavy  $\alpha$ -emitters in nuclear waste and nuclear materials inventories stored in various sites around the world<sup>1</sup>. These include plutonium and minor actinides such as americium and curium. In preparation for geological disposal there is consensus<sup>2</sup> that actinides that have been separated from spent nuclear fuel should be immobilized within mineral-based ceramics rather than glass<sup>2–4</sup> because of their superior aqueous durability and lower risk of accidental criticality. However, in the long term, the  $\alpha$ -decay taking place in these ceramics will severely disrupt their crystalline structure<sup>2–4</sup> and reduce their durability<sup>5,6</sup>. A fundamental property in predicting cumulative radiation damage is the number of atoms permanently displaced per  $\alpha$ -decay. At present, this number is estimated to be 1,000–2,000 atoms/ $\alpha$  in zircon<sup>4</sup>. Here we report nuclear magnetic resonance, spin-counting experiments that measure close to 5,000 atoms/ $\alpha$  in radiation-damaged natural zircons. New radiological nuclear magnetic resonance measurements on highly radioactive, <sup>239</sup>Pu zircon show damage similar to that caused by <sup>238</sup>U and <sup>232</sup>Th in mineral zircons at the same dose, indicating no significant effect of half-life or loading levels (dose rate). On the basis of these measurements, the initially crystalline structure of a 10 weight per cent <sup>239</sup>Pu zircon would be amorphous after only 1,400 years in a geological repository (desired immobilization timescales are of the order of 250,000 years). These measurements establish a basis for assessing the long-term structural durability of actinide-containing ceramics in terms of an atomistic understanding of the fundamental damage event.

Several actinide-containing minerals, such as zircon (ZrSiO<sub>4</sub>), are highly durable, with the result that some are almost as old as the Earth (4.4 gigayears; ref. 7). They have frequently endured extreme geological events while remaining as closed systems for uranium and thorium over geological timescales<sup>8,9</sup>. This long-term actinide retention with respect to a variety of geological events is the major design criterion for ceramic immobilization matrices for the disposal of actinides arising from nuclear power generation or weapons production. Although it may not be the final choice of actinide immobilization matrix (see below), zircon is the best available natural analogue for comparing material behaviour on a geological timescale.

One certainty in the behaviour of actinide-containing ceramic phases over the long term will be the number of  $\alpha$ -decay events they will need to withstand over the proposed lifetime of any geological repository or storage facility. Radiation damage due to the emission of an alpha particle by an actinide nucleus is believed to occur through two distinct processes<sup>4</sup>. The  $\alpha$  particle itself (4.5–5.5 MeV), will mainly cause ionizations during its flight through a material, but it is also known to cause a few hundred atomic displacements (Frenkel defects) as it is slowed by collision with atomic nuclei. The recoil of the heavy actinide nucleus (70–100 keV) is believed to

cause the majority of the localized structural damage (amorphization) as it creates a cascade of collisions with surrounding ions. In minerals, this process is known as metamictization and it was quantifiably linked to actinide  $\alpha$ -activity in zircon 50 years ago<sup>10</sup>. These authors<sup>10</sup> showed that under increasing  $\alpha$ -decay doses of  $\alpha$ -particles and heavy nuclear recoils, severe broadening and shifting of X-ray diffraction peaks accompanied decreases in properties such as the density and birefringence of zircon, evidence of swelling and a loss of crystallinity. Alpha-radiation damage at an atomistic level in bulk samples is correspondingly difficult to follow using diffraction-based techniques. Nuclear magnetic resonance (NMR) spectroscopy, on the other hand, can detect amorphous and crystalline regions equally. Furthermore, the ability of NMR to elucidate local structure in conventionally amorphous materials such as glasses suggests that it can be of similar value in structural studies of the damaged regions<sup>11–13</sup> of crystalline materials.

The effect of increasing  $\alpha$ -radiation damage on the local atomic structure in zircon is illustrated in Fig. 1. The <sup>29</sup>Si NMR spectrum of an undamaged zircon (Fig. 1a) exhibits a narrow line at –81.6 p.p.m. consistent with silicon bonded to four non-bridging oxygens<sup>14</sup> in the zircon structure. In contrast, in a zircon that has received  $1.2 \times 10^{18}$   $\alpha$ /g, but has not exhibited significant swelling (Fig. 1b), the spectrum presents an additional broad resonance, at more negative chemical shifts, that is due to a combination of isolated and polymerized SiO<sub>4</sub> tetrahedra in amorphized material<sup>12,13,15,16</sup>. This indicates the ability of the NMR technique to detect structural damage before it becomes evident in the density. Upon a further increase in the  $\alpha$ -dose, the relative amounts of crystalline and amorphous regions change, and the character of the broad component changes (Fig. 1c) as damaged regions percolate and swelling begins<sup>15,17</sup>. Finally, Fig. 1d shows a high- $\alpha$ -dose zircon with almost complete swelling that presents a substantial broad NMR signal from the majority amorphous phase, although some residual ‘islands’ of crystallinity remain, indicated by the sharper peak.

If the crystalline and amorphous regions of the zircons contribute solely to the narrow and broad components, respectively, of the NMR lineshape (Supplementary Fig. 1), then it is possible to deconvolve the spectra into a crystalline and amorphous fraction  $f_a$ . The direct proportionality of the NMR response to atomic abundance means that it can be used to count the number of atoms (in this case silicon atoms) in the amorphous phase as a function of  $\alpha$ -dose. Figure 2 shows data produced from such an analysis of a series of natural zircons with varying  $\alpha$ -doses. The amorphous fraction here represents the number of silicon atoms in the amorphous phase in contrast to previous determinations of the degree of amorphization in zircon<sup>18</sup> that implicitly report the amorphous volume fraction. The dose is represented here as the number of  $\alpha$ -decays per silicon atom  $D_\alpha$ . This allows us to fit the amorphous fraction  $f_a$  directly to the number

<sup>1</sup>Department of Earth Sciences, University of Cambridge, Downing Street, Cambridge, CB2 3EQ, UK. <sup>2</sup>Pacific Northwest National Laboratory, Richland, Washington 99352, USA.

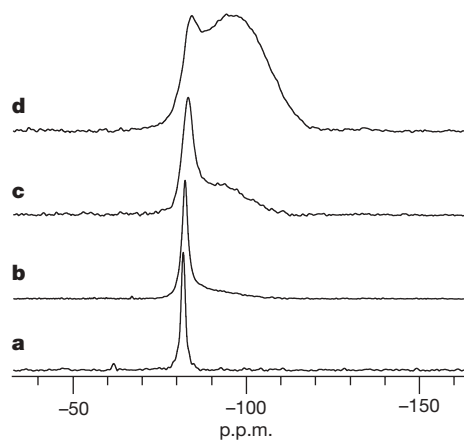


of permanently displaced atoms in each individual  $\alpha$ -decay  $N_d$  via the direct damage equation<sup>19</sup>:

$$f_a = 1 - \exp(-N_d D_\alpha)$$

The fit to data in Fig. 2 gives  $N_d = 830 \pm 50$  silicon atoms, which would correspond to  $4,980 \pm 300$  total atoms displaced if the six-atom  $\text{ZrSiO}_4$  formula unit is used to calculate the total number of atoms displaced ( $6N_d$ ). The total number of displacements is significantly different from existing estimates of 1,000–2000 atomic displacements per  $\alpha$ -event, based solely on Monte Carlo simulations of ballistic (collisional) processes and assumptions of atomic threshold displacement energies<sup>4,20</sup>. This difference is largely associated with local melting in the core of the cascade<sup>21</sup> that is not accounted for in the simple ballistic calculations. Molecular dynamics simulations<sup>22,23</sup> give a better representation of this process and indicate a cascade of  $\sim 5,000$ – $6,000$  atomic displacements for a 30 keV  $^{234}\text{U}$  recoil event in zircon, with physical dimensions of 50–70 Å. The relationship between the number of Si displacements and total displacements is also borne out by detailed analysis in the most recent simulation<sup>23</sup>. Other measures of the size of a cascade arise from single cascade features observed in transmission electron microscopy studies of lightly damaged zircons. These are difficult to detect and quantify<sup>24</sup>, but have been reported to be of the order of 20–50 Å (ref. 21). The volume of 4,980 zircon 'pseudoatoms' would be equal to a sphere 46 Å in diameter.

To extend the spin-counting method of damage detection to potential nuclear waste forms, we have implemented a safe protocol for performing radiological magic-angle spinning (MAS) NMR<sup>25</sup>. We have performed extensive characterization of the relaxation times and  $^{29}\text{Si}$  detection levels and have obtained high signal-to-noise and high-resolution spectra of three plutonium-containing zircons. Figure 3 shows the  $^{29}\text{Si}$  MAS NMR spectrum of  $^{238}\text{Pu}_{0.08}\text{Zr}_{0.92}\text{SiO}_4$  (Fig. 3a) and  $^{239}\text{Pu}_{0.08}\text{Zr}_{0.92}\text{SiO}_4$  (Fig. 3b). These two Pu-containing materials were chemically identical, differing only in the Pu isotope used in their preparation in 1981 (refs 20, 26); in the intervening time, the  $^{239}\text{Pu}$  sample ( $t_{1/2} = 24,100$  years) has received  $1.62 \times 10^{17}$   $\alpha$ /g, whereas the  $^{238}\text{Pu}$  sample ( $t_{1/2} = 87.7$  years) has received  $4.2 \times 10^{19}$   $\alpha$ /g. These spectra exemplify the effect of plutonium  $\alpha$ -decay on the crystal structure of zircon.

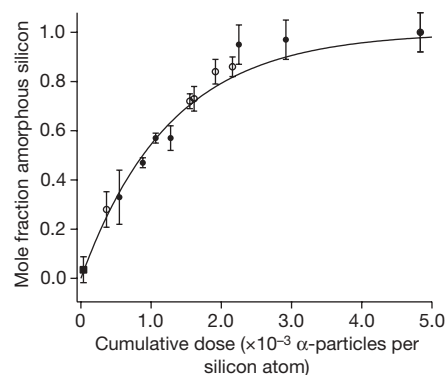


**Figure 1**  $^{29}\text{Si}$  MAS NMR spectra of natural zircons. Increasing cumulative  $\alpha$ -radiation doses cause substantive changes in the local structure around silicon atoms as the amorphization of the crystals progresses. **a**, An undamaged zircon. **b**, UG40, a zircon with  $1.2 \times 10^{18}$   $\alpha$ /g. **c**, Cam26, a zircon with  $2.9 \times 10^{18}$   $\alpha$ /g. **d**, Ni12, a zircon with  $7.1 \times 10^{18}$   $\alpha$ /g. In each case, the sharper peak in the spectrum arises from silicon atoms in crystalline domains and the broader peak at lower frequency (more negative p.p.m. values) arises from silicon atoms in amorphized regions. Relative areas are obtained by curve-fitting the crystalline peak and subtracting this intensity from the integrated intensity of the whole lineshape.

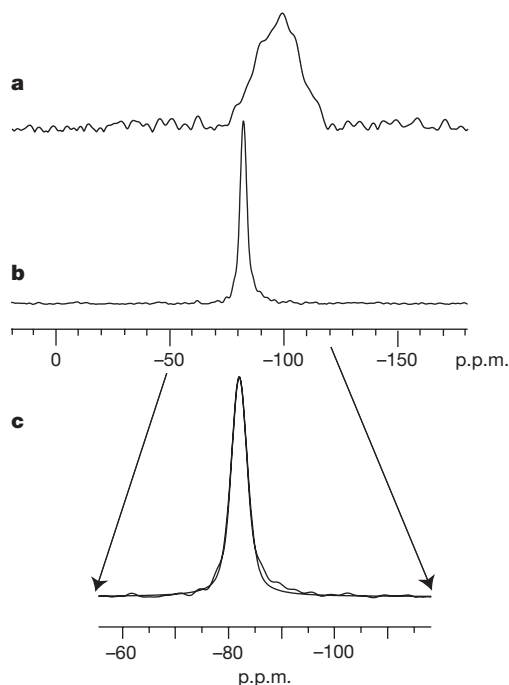
The spectrum in Fig. 3a of the  $^{238}\text{Pu}$ -doped zircon exhibits a broad line centred at  $-98$  p.p.m., typical of a fully amorphous silicate material, and no narrow peak in the  $-81$  to  $-84$  p.p.m. region characteristic of natural crystalline zircons. This result agrees with data from natural zircons with doses of greater than  $1.2 \times 10^{19}$   $\alpha$ /g that present as fully amorphous. The  $^{29}\text{Si}$  NMR spectrum of  $^{239}\text{Pu}_{0.08}\text{Zr}_{0.92}\text{SiO}_4$  is shown in Fig. 3b. The resonance position of this peak is identical to that in natural zircon, indicating that there is no transferred hyperfine field from the potentially unpaired plutonium 5f electrons that could shift the Si resonance frequency for those Si adjacent to Pu substitution sites in the zircon crystal structure. A random distribution of Pu substitution onto Zr sites would position 32% of Si atoms next to at least one Pu atom. There is a small amount of detectable damage to more negative chemical shifts in this  $^{239}\text{Pu}$ -doped sample. The expanded region in Fig. 3c shows the superposition of a fitted crystalline component and the total lineshape. This reveals an area of  $3 \pm 2\%$  of the total lineshape attributed to amorphized material, which is plotted on the damage curve for natural zircons in Fig. 2.

The results of Si signal detection tests on this ceramic sample show that we detect all of the silicon atoms in the sample within a precision of 5% imposed by the sample mass measurements and the signal-to-noise limits of single pulse acquisitions with long  $^{29}\text{Si}$  relaxation times. The value of  $f_a$  predicted for the cumulative damage in  $^{239}\text{Pu}$  ceramic zircon over 25 years would be 0.04, if  $N_d$ , the damage per  $\alpha$ -event derived from the  $^{238}\text{U}$  (and  $^{232}\text{Th}$ ) in natural zircons, is used. In principle, differences in  $\alpha$ -induced structural changes between natural ( $^{238}\text{U}$ ) and synthetic ( $^{239}\text{Pu}$ ) zircons could be influenced by the increased dose rate arising from the shorter half-life of  $^{239}\text{Pu}$  and the increased loading: 8 mol% versus  $\sim 1$  mol%. The agreement in the damage fraction for  $^{239}\text{Pu}$  and  $^{238}\text{U}$  zircons is evidence of the lack of a dose rate effect.

The fundamental damage fraction measured here allows the further projection of damage accumulation in  $^{239}\text{Pu}$  containing zircons on a repository lifetime scale. This would indicate that a 10 wt% Pu-loaded zircon wastef orm ( $\text{Pu}_{0.08}\text{Zr}_{0.92}\text{SiO}_4$ ) would pass the percolation limit<sup>27</sup> and start to swell significantly after 210 years; it would be completely ( $>99\%$ ) amorphous after 1,400 years. This time is very short in terms of the ideal immobilization of  $^{239}\text{Pu}$  for ten half-lives or 241,000 years. Ceramic actinide matrices will more likely be



**Figure 2** Amorphous mole fraction of silicon as a function of accumulated  $\alpha$ -dose in natural zircons. The accumulated dose is expressed as the number of  $\alpha$ -particles accumulated per silicon atom. The solid circles are data from a previous study<sup>11</sup> acquired with single pulse excitation, while new data on small samples (open circles) and a  $^{239}\text{Pu}_{0.08}\text{Zr}_{0.92}\text{SiO}_4$  specimen (square) were obtained with multiple echo techniques. Error bars indicate the precision of the relative amounts of silicon in crystalline and amorphous regions based on the detection level of the  $^{29}\text{Si}$  signal. They indicate the bounds if all undetected signal resided uniquely in a crystalline or an amorphous region. In the case of complete detection, the error bar indicates the range of values obtained by varying the initial parameters in a least-squares fitting procedure.



**Figure 3 | The effect of plutonium  $\alpha$ -self-irradiation on the local structure of Pu-doped ceramic zircons.**  $^{29}\text{Si}$  MAS NMR spectra of  $^{238}\text{Pu}_{0.08}\text{Zr}_{0.92}\text{SiO}_4$  (a) and  $^{239}\text{Pu}_{0.08}\text{Zr}_{0.92}\text{SiO}_4$  (b). The expansion in c of this high-resolution and high-sensitivity spectrum reveals the onset of internal radiation damage in the  $^{239}\text{Pu}$  material. A damage fraction of  $0.03 \pm 0.02$  can be computed by subtracting the fitted crystalline component.

titanate or zirconate pyrochlore<sup>28</sup> or structurally related zirconolite materials, which can accommodate additional radionuclide species in multiphase ceramics. The application of techniques such as those described here, that are sensitive to the local effects of  $\alpha$ -radiation damage and its quantification in the bulk, will substantially improve long-term confidence in these materials as actinide immobilization matrices.

## METHODS

**NMR acquisition protocol.** The  $^{29}\text{Si}$  nuclear magnetic resonance spectra presented here were acquired using rotor-synchronized Carr–Purcell–Meiboom–Gill (CPMG) echo techniques<sup>29</sup>. When  $T_1$  and  $T_2$ —the longitudinal spin-lattice and transverse relaxation times, respectively—are both long, these techniques improve the signal-to-noise ratio significantly compared with single pulse acquisitions<sup>15</sup>. The approach also achieves a higher Si detection level when compared with small-pulse-angle, single-pulse acquisitions used previously<sup>15</sup>. Here the spins were polarized for 3,600 s and approached full magnetization (>95% for natural zircons, 100% for synthetic zircons). Each polarization period is followed by the acquisition of a series of 32 or 64 spin-echoes. These echoes are separated and then co-added to produce a ‘whole echo’ with very good signal-to-noise ratio. This whole echo is apodized and Fourier-transformed in magnitude mode so that no phase correction is required. This methodology permits the examination of small sample masses, in the range 20–50 mg, thus enabling a more complete assessment of the  $\alpha$ -dose versus amorphous fraction ( $f_a$ ) curve for natural zircons by using some of the well-characterized samples used in earlier diffraction and spectroscopy studies<sup>17,30</sup>. The acquisition of spectra of relatively small amounts of highly radioactive synthetic zircons containing plutonium is also enabled. The spectrum in Fig. 4b was the sum of 16 CPMG acquisitions consisting of 64 echoes separated by 25.6 ms with a polarization delay of 3,500 s. The resolution of the spectra depends on the echo separation: the larger the echo separation, the longer the signal acquisition in the time domain and the higher is the resolution in the frequency domain. For the natural zircons, echo separations of 25.6 ms were used to give a digital resolution of 78 Hz, which is of the order of the linewidth (1 p.p.m.) in single-pulse acquisitions of undamaged zircons. Similarly, the linewidth of the  $^{239}\text{Pu}_{0.08}\text{Zr}_{0.92}\text{SiO}_4$  spectrum did not decrease when the individual echo acquisition time was increased from 12.5 to 25.6 ms, indicating that a natural limit had been reached.

**Radiological MAS NMR.** Because  $^{239}\text{Pu}$  and  $^{238}\text{Pu}$  are  $\alpha$ -emitters that present considerable radiotoxicity hazards, all sample manipulations involving plutonium-containing ceramics, including NMR measurements, were performed by supervised, trained personnel in the Radiochemical Processing Laboratory, which is a US Department of Energy Category 2 nuclear facility at the Pacific Northwest National Laboratory. Three samples of weighed ceramic cores of zircon were sealed into triple containment according to a previously published protocol<sup>25</sup> for MAS NMR. These samples had masses of 157.5 mg ( $^{239}\text{Pu}_{0.04}\text{Zr}_{0.92}\text{SiO}_4$ ), 138.2 mg ( $^{239}\text{Pu}_{0.08}\text{Zr}_{0.92}\text{SiO}_4$ ) and 159.3 mg ( $^{238}\text{Pu}_{0.08}\text{Zr}_{0.92}\text{SiO}_4$ ) with specific activities of  $1.15 \times 10^8 \text{ Bq g}^{-1}$ ,  $2.30 \times 10^8 \text{ Bq g}^{-1}$  and  $4.46 \times 10^{10} \text{ Bq g}^{-1}$ , respectively. The NMR spectra of radioactive samples were recorded at a Larmor frequency of 59.64 MHz (7.0 T) on a Tecmag Discovery 300 spectrometer and a spinning speed of 3.5 kHz. Measurement of the Si detection level in these samples was accomplished by acquisition of the spectra of a weighed standard of tetrakis(tri-methyl)silane (TKS) under identical conditions of triple containment, tuning and matching and the signals compared on a per atom per scan basis for the acquisition of  $90^\circ$  pulses separated by delays of  $5T_1$  (Supplementary Information, Fig. 2). All  $^{29}\text{Si}$  spectra were referenced to tetramethylsilane (TMS) via a secondary reference of silicone rubber.

Received 30 June; accepted 3 November 2006.

1. International Atomic Energy Agency. (IAEA). *Guidelines for the Management of Plutonium (INFCIRC/549): Background and Declarations* ([www.isis-online.org/global\\_stocks/old/guidelines\\_for\\_management.html](http://www.isis-online.org/global_stocks/old/guidelines_for_management.html)) (Institute for Science and National Security, 2003).
2. Muller, I. & Weber, W. J. Plutonium in crystalline ceramics and glasses. *MRS Bull.* **26**, 698–706 (2001).
3. Weber, W. J. *et al.* Radiation effects in glasses used for immobilization of high-level waste and plutonium disposition. *J. Mater. Res.* **12**, 1946–1978 (1997).
4. Weber, W. J. *et al.* Radiation effects in crystalline ceramics for the immobilization of high-level nuclear waste and plutonium. *J. Mater. Res.* **13**, 1434–1484 (1998).
5. Ewing, R. C. Nuclear waste forms for actinides. *Proc. Natl Acad. Sci. USA* **96**, 3432–3439 (1999).
6. Balan, E. *et al.* Metamictization and chemical durability of detrital zircon. *Am. Mineral.* **86**, 1025–1033 (2001).
7. Wilde, S. A., Valley, J. W., Peck, W. H. & Graham, C. M. Evidence from detrital zircons for the existence of continental crust and oceans on the Earth 4.4 Gyr ago. *Nature* **409**, 175–178 (2001).
8. Ewing, R. C. The design and evaluation of nuclear-waste forms: Clues from mineralogy. *Can. Mineral.* **39**, 697–715 (2001).
9. Ewing, R. C., Meldrum, A., Wang, L. M., Weber, W. J. & Corrales, L. R. Radiation effects in zircon. *Rev. Mineral. Geochem.* **53**, 387–425 (2003).
10. Holland, H. D. & Gottfried, D. The effect of nuclear radiation on the structure of zircon. *Acta Crystallogr.* **8**, 291–300 (1955).
11. Farnan, I., Balan, E., Pickard, C. J. & Mauri, F. The effect of radiation damage on local structure in the crystalline fraction of  $\text{ZrSiO}_4$ : Investigating the  $^{29}\text{Si}$  NMR response to pressure in zircon and reidite. *Am. Mineral.* **88**, 1663–1667 (2003).
12. Balan, E., Mauri, F., Pickard, C. J., Farnan, I. & Calas, G. The aperiodic states of zircon: an ab initio molecular dynamics study. *Am. Mineral.* **88**, 1769–1777 (2003).
13. Ashbrook, S. E. & Farnan, I. Solid-state  $^{17}\text{O}$  nuclear magnetic resonance spectroscopy without isotopic enrichment: direct detection of bridging oxygen in radiation damaged zircon. *Solid State Nucl. Magn. Reson.* **26**, 105–112 (2004).
14. Magi, M., Lippmaa, E., Samoson, A., Engelhardt, G. & Grimmer, A. R. Solid-state high resolution  $^{29}\text{Si}$  chemical shifts in silicates. *J. Phys. Chem.* **88**, 1518–1522 (1984).
15. Farnan, I. & Salje, E. K. H. The degree and nature of radiation damage in zircon observed by  $^{29}\text{Si}$  nuclear magnetic resonance. *J. Appl. Phys.* **89**, 2084–2090 (2001).
16. Farnan, I.  $^{29}\text{Si}$  NMR characterisation of the crystalline-amorphous transition in  $\text{ZrSiO}_4$ . *Phase Transit.* **69**, 47–60 (1999).
17. Rios, S., Salje, E. K. H., Zhang, M. & Ewing, R. C. Amorphization in zircon: evidence for direct impact damage. *J. Phys. Condens. Matt.* **12**, 2401–2412 (2000).
18. Murakami, T., Chakoumakos, B. C., Ewing, R. C., Lumpkin, G. R. & Weber, W. J. Alpha-decay event damage in zircon. *Am. Mineral.* **76**, 1510–1532 (1991).
19. Gibbons, J. F. Ion implantation in semiconductors. 2. Damage production and annealing. *Proc. IEEE* **60**, 1062–1096 (1972).
20. Weber, W. J. Alpha-decay-induced amorphization in complex silicate structures. *J. Am. Ceram. Soc.* **76**, 1729–1738 (1993).
21. Meldrum, A., Zinkle, S. J., Boatner, L. A. & Ewing, R. C. A transient liquid-like phase in the displacement cascades of zircon, hafnon and thorite. *Nature* **395**, 56–58 (1998).
22. Trachenko, K., Dove, M. T. & Salje, E. K. H. Structural changes in zircon under alpha-decay irradiation. *Phys. Rev. B* **65**, 180102 (2002).
23. Devanathan, R. *et al.* Molecular dynamics simulation of energetic recoil damage in zircon. *Molec. Simul.* (in the press).



24. Miller, M. L. & Ewing, R. C. Image simulation of partially amorphous materials. *Ultramicroscopy* **48**, 203–237 (1992).
25. Farnan, I. *et al.* High-resolution solid-state nuclear magnetic resonance experiments on highly radioactive ceramics. *Rev. Sci. Instrum.* **75**, 5232–5236 (2004).
26. Weber, W. J. Self-radiation damage and recovery in Pu-doped zircon. *Radiat. Effects Defects Solids* **115**, 341–349 (1991).
27. Trachenko, K., Dove, M. T. & Salje, E. K. H. Large swelling and percolation in irradiated zircon. *J. Phys. Cond. Matt.* **15**, L1–L7 (2003).
28. Ewing, R. C., Weber, W. J. & Lian, J. Nuclear waste disposal-pyrochlore (A<sub>2</sub>B(2)O(7)): Nuclear waste form for the immobilization of plutonium and “minor” actinides. *J. Appl. Phys.* **95**, 5949–5971 (2004).
29. Larsen, F. H. & Farnan, I. <sup>29</sup>Si and <sup>17</sup>O (Q)CPMG-MAS solid-state NMR experiments as an optimum approach for half-integer nuclei having long T-1 relaxation times. *Chem. Phys. Lett.* **357**, 403–408 (2002).
30. Zhang, M. *et al.* Metamictization of zircon: Raman spectroscopic study. *J. Phys. Cond. Matt.* **12**, 1915–1925 (2000).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank N. R. Johnson, A. E. Kozelisky and R. D. Scheele for help with initial experimental development, sample handling and logistical support, and M. Zhang for help with the natural zircons. I.F. acknowledges funding from the UK EPSRC. A portion of the research described here was performed under a user programme at the Environmental Molecular Sciences Laboratory of the Pacific Northwest National Laboratory. PNNL staff and work were supported by the Environmental Management Science Program, Office of Biological and Environmental Research, US Department of Energy.

**Author Contributions** I.F. carried out the NMR work on natural zircons. I.F. conceived the radioactive MAS NMR experiments and I.F. and H.C. developed the technology and performed them in H.C.’s laboratory. W.J.W. provided the Pu zircon ceramics, supporting characterization data and radiological support. I.F. wrote the paper and all authors had a chance to contribute to and comment on the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to I.F. ([ifarnan@esc.cam.ac.uk](mailto:ifarnan@esc.cam.ac.uk)).

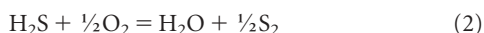
## LETTERS

## Redox evolution of a degassing magma rising to the surface

Alain Burgisser<sup>1</sup> & Bruno Scaillet<sup>1</sup>

Volatiles carried by magmas, either dissolved or exsolved, have a fundamental effect on a variety of geological phenomena, such as magma dynamics<sup>1–5</sup> and the composition of the Earth's atmosphere<sup>6</sup>. In particular, the redox state of volcanic gases emanating at the Earth's surface is widely believed to mirror that of the magma source, and is thought to have exerted a first-order control on the secular evolution of atmospheric oxygen<sup>6,7</sup>. Oxygen fugacity ( $f_{O_2}$ ) estimated from lava or related gas chemistry, however, may vary by as much as one log unit<sup>8–10</sup>, and the reason for such differences remains obscure. Here we use a coupled chemical–physical model of conduit flow to show that the redox state evolution of an ascending magma, and thus of its coexisting gas phase, is strongly dependent on both the composition and the amount of gas in the reservoir. Magmas with no sulphur show a systematic  $f_{O_2}$  increase during ascent, by as much as 2 log units. Magmas with sulphur show also a change of redox state during ascent, but the direction of change depends on the initial  $f_{O_2}$  in the reservoir. Our calculations closely reproduce the  $H_2S/SO_2$  ratios of volcanic gases observed at convergent settings, yet the difference between  $f_{O_2}$  in the reservoir and that at the exit of the volcanic conduit may be as much as 1.5 log units. Thus, the redox state of erupted magmas is not necessarily a good proxy of the redox state of the gases they emit. Our findings may require re-evaluation of models aimed at quantifying the role of magmatic volatiles in geological processes.

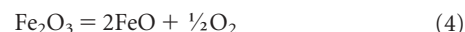
Recent studies have investigated the role of redox equilibria during volatile exsolution, but the solution of the numerical problem requires fixing either iron or sulphur redox states<sup>11</sup>: as a result, the effect of decompression on  $f_{O_2}$  cannot be evaluated. In the present work, we relax such an assumption by taking advantage of the fact that the solubility laws of key volatile species other than  $H_2O$  and  $CO_2$  have been recently determined, in particular those of  $H_2$ ,  $SO_2$  and  $H_2S$  (refs 12, 13). We consider gas phases in the system H–O–S, with six species ( $H_2O$ ,  $H_2$ ,  $SO_2$ ,  $H_2S$ ,  $S_2$  and  $O_2$ ), in which the following redox equilibria occur:



Standard thermodynamic considerations<sup>14,15</sup> show that once total pressure ( $P$ ), temperature ( $T$ ), and two additional intensive parameters such as water and hydrogen fugacities ( $f_{H_2O}$  and  $f_{H_2}$ ) are known, the fugacities of all remaining species ( $f_{H_2S}$ ,  $f_{O_2}$ ,  $f_{S_2}$ ,  $f_{SO_2}$ ) are fixed and the gas phase composition in the H–O–S system is fully determined. Each species  $i$  in the gas has a mole fraction  $m_i$ ;  $\sum m_i = 1$ .

The gas phase is modelled as an ideal mixture of non-ideal gases, a valid approximation in the pressure range considered here. Departure from ideal behaviour of end-member species is accounted for by the fugacity coefficient  $\gamma_i$ , which is fixed by  $P$  and  $T$  (ref. 16). The total weight fraction of each species ( $w_{Ti}$ ) is the sum of its exsolved part ( $w_{gi}$ ) and its dissolved part:  $w_{Ti} = w_{gi} + a_i(\gamma_i m_i P)^{b_i}$ , where  $a_i$  and  $b_i$  are experimentally determined solubility constants (Table 1). We use a homogeneous, one-dimensional conduit flow model<sup>5</sup> to simulate magma ascent under closed-system and equilibrium conditions. Magma rises in a cylindrical conduit at constant mass flux, and volatile exsolution affects the flow through changes in buoyancy and viscosity (see Supplementary Information). Calculations are performed by first fixing  $P$ ,  $T$ ,  $f_{H_2O}$ ,  $f_{H_2}$  and the amount of gas in the reservoir. The model seeks first the equilibrium distribution of each volatile species between gas and melt before ascent. Then, at each next lower pressure, mass conservation requires that total amounts of each element (O, H and S) remain constant. Using this constraint, the model calculates the equilibrium distribution of volatile species, which in turn affects ascent dynamics.

Redox equilibrium during magma ascent involving dissolved iron can be written as:



The importance of such a reaction will be dictated by the initial amount of dissolved iron oxides. We have simulated the redox effect of iron by using an empirical model that relates the  $FeO/Fe_2O_3$  ratio of silicate melts to  $f_{O_2}$  (ref. 17); here  $f_{O_2}$  is referenced to the solid buffer Ni–NiO, such that  $NNO + 1$  means an  $f_{O_2}$  one order of magnitude higher than  $NNO$ ). Runs under typical storage conditions of arc rhyolites ( $\leq 1$  wt% total iron and  $f_{O_2}$  between  $NNO - 1$  and  $NNO + 1$ ) show that reaction (4) partly buffers changes in  $f_{O_2}$  when little gas is present in the reservoir ( $< 0.2$  log units without changing the redox trend during ascent, see Supplementary Information). In contrast, in iron-rich liquids such as basalts, the buffering capacity of iron species will be higher. Thus our results primarily apply to magmas in which the residual melt is rhyolitic, as commonly observed in arc settings. We did not consider the role of iron in crystals because

**Table 1 | Solubility constants**

Species	$a_i$	$b_i$
$H_2O$	$1.063 \times 10^{-3}$	0.5399
$H_2$	$3.400 \times 10^{-7}$	1.2800
$SO_2$	$1.632 \times 10^{-10}$	1.3789
$H_2S$	$8.239 \times 10^{-6}$	0.5145
$O_2$	0	0
$S_2$	0	0

The parameters  $a_i$  and  $b_i$  were determined by fitting experimental solubility data of corresponding species to an empirical equation of the form  $w_i = a_i f_i^{b_i}$ . We have used data from ref. 12 for sulphur-bearing species, from ref. 13 for hydrogen, and from ref. 29 for water.

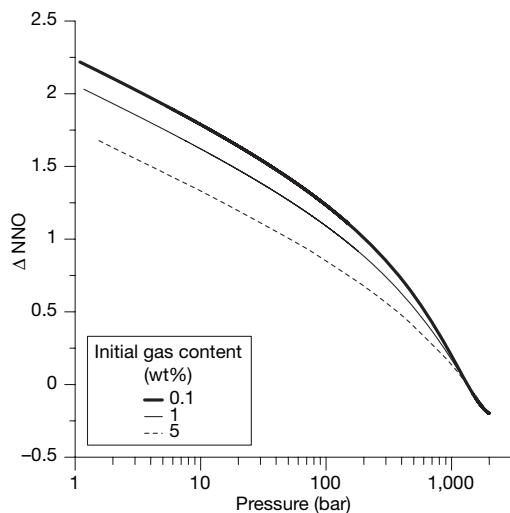
<sup>1</sup>ISTO, UMR 6113 Université d'Orléans-CNRS, 1a rue de la Férollerie, 45071 Orléans cedex 2, France.



the role of solid buffers as a first-order mechanism controlling redox state during magma ascent can be ruled out on kinetic grounds. The lattice diffusion processes that this mechanism requires are exceedingly slow compared to those in gas or liquid phases.

Similarly, for simplicity we have not investigated the role of  $\text{CO}_2$ , as detailed petrologic studies of silicic magmas in arcs show them to have little or no  $\text{CO}_2$  under pre-eruptive conditions<sup>18</sup>. In general, however, introduction of  $\text{CO}_2$  will lower  $f_{\text{H}_2\text{O}}$  and thus drive our calculated values towards lower  $f_{\text{O}_2}$  via equilibrium (1). Our assumption of equilibrium implies that the model might not capture the chemical evolution of the gas within rapidly decompressed magmas, such as in plinian eruptions, in which the contrasted diffusive kinetics of volatile species may inhibit attainment of equilibrium. Both  $\text{H}_2\text{O}$  and  $\text{H}_2$  are, however, fast diffusing species relative to  $\text{CO}_2$  and S-bearing ones<sup>18,19</sup>. Thus, if physical fractionation of volatile species arises during ascent, the gas phase composition will be driven towards the system H–O (C and S species remain in the melt, as documented for S for the 1991 Pinatubo eruption<sup>20</sup>), which will exert a dominant control on the redox state of escaping gases. Our model thus represents a fundamental end-member case towards which magmas, notably those andesitic to silicic in arc settings, tend to evolve.

We have explored the following range of starting conditions, typical of silicic arc magmas<sup>21,22</sup>: an initial pressure from 2,000 to 3,000 bar,  $f_{\text{O}_2}$  from NNO + 2 to NNO – 0.5, bulk iron contents ( $\text{FeO}^* = \text{FeO} + \text{Fe}_2\text{O}_3/1.113$ ) up to 3 wt%, bulk water contents up to 10.4 wt% and bulk sulphur contents up to 3 wt%, the last two parameters being adjusted by varying the amount of excess gas in the reservoir (up to 5 wt%; ref. 23). The conduit radius was fixed at 5 m for all runs. Temperature has been fixed to 825 °C, that is, typical of rhyolite magmas<sup>21</sup>, and melt density to 2,140 kg m<sup>–3</sup>, though different choices will not affect the trends observed. Runs are constrained to reach atmospheric pressure at the vent, which yields initial ascent rates between 0.7 and 12 m s<sup>–1</sup>. Because degassing occurs in equilibrium, changes in chemistry as a function of pressure occur regardless of ascent speed. In all cases, the simulations are carried down to atmospheric pressure, though it can be anticipated that the last increments of  $f_{\text{O}_2}$  change we compute might not be



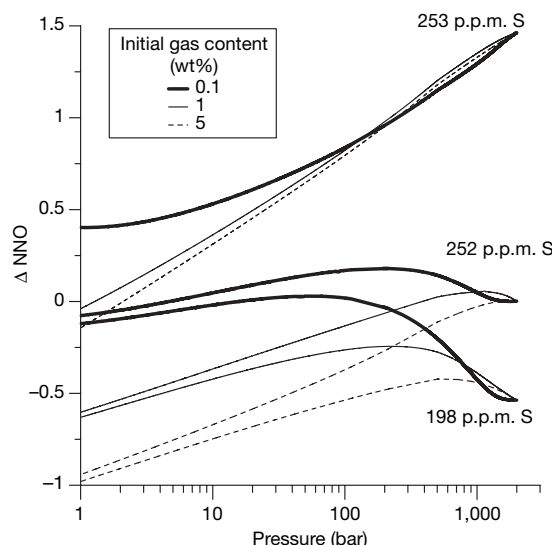
**Figure 1 | Fundamental relationship between magma ascent and magma redox state, for a rhyolite magma coexisting with a H–O gas.** From gas content and composition at depth, the coupled model of conduit flow calculates the evolution of the physical and chemical conditions of the ascending magma. The effect of equilibrium degassing is a systematic increase in  $f_{\text{O}_2}$  with decreasing pressure (that is, with shallower depth). Here  $f_{\text{O}_2}$  is referenced to the solid buffer Ni–NiO (NNO); NNO + 1 means an  $f_{\text{O}_2}$  one order of magnitude higher than NNO, and is expressed as  $\Delta\text{NNO} = 1$ . The figure shows representative cases, for three initial values of exsolved volatiles with an  $f_{\text{H}_2}$  initially fixed at 10 bar.

reached, owing either to the inhibiting effect of viscosity at low water content<sup>24</sup> or to gas loss once a permeability threshold is reached such that the system becomes open to gas<sup>25</sup>.

We first consider the S- and Fe-free case. All simulations performed showed essentially identical behaviour in terms of redox state evolution, that is, the  $f_{\text{O}_2}$  of the magma increases during ascent, the magnitude of increase being more or less damped by the amount of free gas present in the reservoir. A representative example is shown in Fig. 1, corresponding to a magma initially stored at 2,000 bar, at an  $f_{\text{O}_2}$  of NNO – 0.2, with a melt  $\text{H}_2\text{O}$  content of 5.46 wt%, and for excess gas contents of 0.1, 1 and 5 wt%. The simulations show that the magnitude of  $f_{\text{O}_2}$  change increases when the amount of excess gas decreases, and can reach more than 2 log units for a gas-poor magma (0.1 wt%).

We now explore the case of S-bearing rhyolite magma with 1 wt% of total iron and stored at 2,000 bar, illustrating our calculations with three different initial  $f_{\text{O}_2}$  but similar initial dissolved S contents (~200–250 p.p.m.). A magma starting at an  $f_{\text{O}_2}$  of NNO – 0.5 displays a continuous increase in its  $f_{\text{O}_2}$  as it ascends, except in the last few hundred bars where a reversal in  $f_{\text{O}_2}$  towards reduction occurs (Fig. 2). The magnitude of change is strongly dependent on the amount of gas initially present in the reservoir. At low gas content (0.1 wt%), the  $f_{\text{O}_2}$  rises by 0.7 log units relative to starting conditions. With 5 wt% gas, the magma has a redox state nearly constant up to a pressure of 100 bar. When the starting  $f_{\text{O}_2}$  in the reservoir is at NNO (Fig. 2), the magma undergoes oxidation only for gas-poor conditions (0.1 wt%). Higher amounts of gas in the reservoir impart a reducing trend in the  $f_{\text{O}_2}$  evolution during ascent, the final  $f_{\text{O}_2}$  differing by almost 1 log unit from the initial value for an initial gas content of 5 wt%. When the initial  $f_{\text{O}_2}$  in the reservoir is at NNO + 1.5, the magma undergoes a significant reduction during ascent regardless of its initial gas content (Fig. 2). In this case, the drop in  $f_{\text{O}_2}$  may exceed 1.5 log units at near-atmospheric conditions for gas-rich conditions.

The change in  $f_{\text{O}_2}$  during decompression is accompanied by dramatic changes in gas phase composition (Fig. 3). Our calculated  $\text{H}_2\text{S}/\text{SO}_2$  ratios fall in the range 0.1–10, which is comparable to that of volcanic gases measured at convergent settings for silicic to intermediate magmas<sup>26</sup> (Fig. 3). Clearly, a variety of  $\text{H}_2\text{S}/\text{SO}_2$  ratios can be



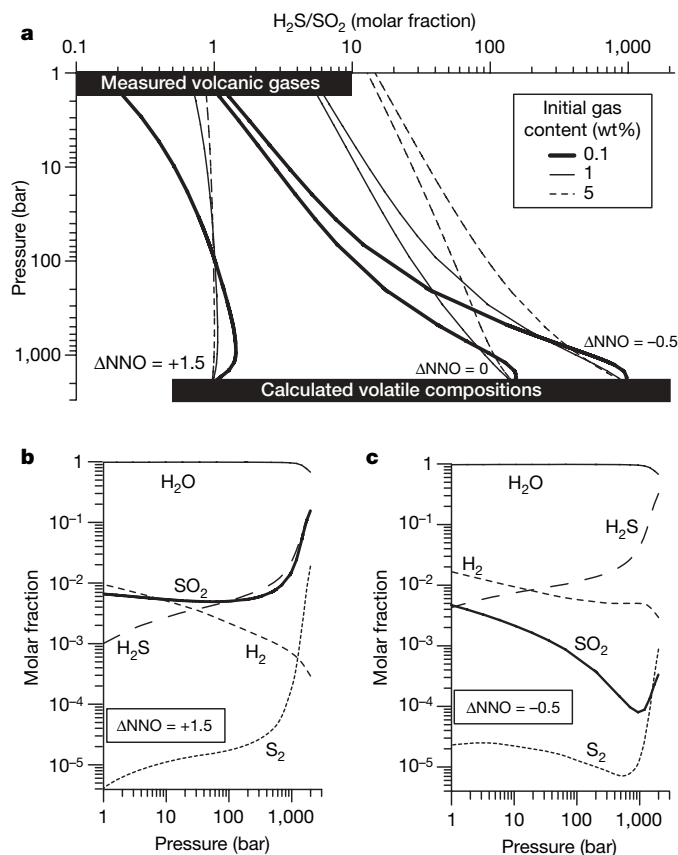
**Figure 2 | Fundamental relationship between magma ascent and magma redox state, for a rhyolite magma coexisting with a H–O–S gas.** The effect of initial redox state on its evolution during decompression is shown. The graph shows representative cases, for three initial values of exsolved volatiles with an  $f_{\text{H}_2\text{O}}$  initially fixed at 1,000 bar. The initial redox state was achieved by varying  $f_{\text{H}_2}$  ( $\Delta\text{NNO} = -0.5$ ,  $f_{\text{H}_2} = 10$ ;  $\Delta\text{NNO} = 0$ ,  $f_{\text{H}_2} = 5.38$ ;  $\Delta\text{NNO} = +1.5$ ,  $f_{\text{H}_2} = 1$ ).

produced from magmas having common initial redox states but different amounts of gases. Conversely, a given  $\text{H}_2\text{S}/\text{SO}_2$  ratio may be produced from a wide range of starting redox conditions. For instance, a  $\text{H}_2\text{S}/\text{SO}_2$  ratio of about one can be produced from a magma initially stored at  $\text{NNO} + 1.5$  with 5 wt% gas, or from a magma initially at  $\text{NNO} - 0.5$  with 0.1 wt% gas (Fig. 3a). Thus, anticipating the  $\text{H}_2\text{S}/\text{SO}_2$  ratio of gases emanating from a given reservoir would require knowing not only the reservoir  $f_{\text{O}_2}$ , but also the amount of free gas present in the reservoir and the depth at which gas and melt are physically separated. Although such a rich behaviour precludes a simple explanation for each trend calculated, a sensitivity analysis of our model suggests that water exsolution plays an important role in oxidizing the system, and that, on the other hand, the complex pressure dependence of the redox equilibria (1)–(3) contribute to the reducing trends. Our findings have thus obvious implications for the use of volcanic gases as a monitoring tool of volcanic activity. They also illustrate the important role of sulphur. Magmas poor in sulphur or in which reactions involving sulphur are kinetically inhibited are likely to undergo a significant increase in their redox state during ascent.

The above results show that the redox state that magma records at depth does not necessarily mirror that of its escaping gases, in particular when they are released from levels shallower than the main reservoir. Thus, from a broader perspective, our findings have implications for our understanding of how past volcanic activity may have affected Earth's atmosphere. Current models of the evolution of

atmospheric oxygen implicitly assume that the redox state of magmatic rocks can be taken as equal to that of their outgassed products<sup>6,7</sup>. Our calculations show that, for silicic magmas, this assumption holds true only under a restricted set of conditions (for instance, a magma starting at  $\text{NNO} - 0.5$  with 1 wt% gas, Fig. 2). Oxidized silicic magmas are particularly prone to redox change during ascent. A recent study<sup>10</sup> has stressed the difference in  $f_{\text{O}_2}$  retrieved from volcanic gas<sup>11</sup> and volcanic glass<sup>12</sup> at Kilauea volcano: the basaltic glass, which is fully degassed, records an  $f_{\text{O}_2}$  that is 1.2 log units lower than that of the gas. Although our model is calibrated on Fe-poor liquids, its predictions are qualitatively in accord with such an observation, which suggests that even for basaltic magmas in non-arc settings, redox change during degassing may occur<sup>27</sup>. The corollary is that the iron redox state of a magma may differ significantly from that of its source, in contrast to conventional wisdom<sup>28</sup>. Altogether, this suggests that equating the redox state and composition of present-day volcanic gases to those emitted in the geologic past<sup>6</sup> may not be a correct assumption.

Received 3 April; accepted 28 November 2006.



**Figure 3 | Evolution of the composition of an H–O–S gas during ascent of a rhyolite magma.** **a**, Effect of the initial redox state on the ratio  $\text{H}_2\text{S}/\text{SO}_2$  for three starting values of  $f_{\text{O}_2}$ , each with three different gas contents (same initial conditions as in Fig. 2). Also shown are the natural range observed on active volcanoes in convergent settings<sup>26</sup>, and the range calculated by phase equilibria experiments<sup>21</sup>. **b**, Evolution of the gas composition for a magma oxidized at depth ( $\Delta\text{NNO} = +1.5$ , 0.1 wt% gas). **c**, Evolution of the gas composition for a magma reduced at depth ( $\Delta\text{NNO} = -0.5$ , 0.1 wt% gas). Contents of  $\text{O}_2$  are too low (<10 p.p.m.) to be displayed.

- Wilson, L., Sparks, R. S. J. & Walker, G. P. L. Explosive volcanic eruptions. IV. The control of magma properties and conduit geometry on eruption column behavior. *Geophys. J. R. Astron. Soc.* **63**, 117–148 (1980).
- Papale, P. Strain-induced magma fragmentation in explosive eruptions. *Nature* **397**, 425–428 (1999).
- Huppert, H. E. & Woods, A. W. The role of volatiles in magma chamber dynamics. *Nature* **420**, 493–495 (2002).
- Gonnerman, H. M. & Manga, M. Explosive volcanism may not be an inevitable consequence of magma fragmentation. *Nature* **426**, 432–435 (2003).
- Burgisser, A. & Gardner, J. Experimental constraints on degassing and permeability in volcanic conduit flow. *Bull. Volc.* **67**, 42–56 (2005).
- Holland, H. D. Volcanic gases, black smokers and the great oxidation event. *Geochim. Cosmochim. Acta* **66**, 3811–3826 (2002).
- Kasting, J. F., Egger, D. H. & Raeburn, S. P. Mantle redox evolution and the oxidation state of the atmosphere. *J. Geol.* **101**, 245–257 (1993).
- Gerlach, T. M. Comment on paper 'Morphology and compositions of spinel in Pu'u'O'o lava (1996–1998)', Kilauea volcano, Hawaii—enigmatic discrepancies between lava and gas-based  $f_{\text{O}_2}$  determinations of Pu'u'O'o lava. *J. Volcanol. Geotherm. Res.* **134**, 241–244 (2004).
- Gerlach, T. M. Oxygen buffering of Kilauea volcanic gases and the oxygen fugacity of Kilauea basalt. *Geochim. Cosmochim. Acta* **57**, 795–814 (1993).
- Roeder, P. L., Thornber, C., Proustovetov, A. & Grant, A. Morphology and composition of spinel in Pu'u'O'o lava (1996–1998), Kilauea volcano, Hawaii. *J. Volcanol. Geotherm. Res.* **123**, 245–265 (2003).
- Moretti, R. & Papale, P. On the oxidation state and volatile behavior in multicomponent gas–melt equilibria. *Chem. Geol.* **213**, 265–280 (2004).
- Clemente, B., Scaillet, B. & Pichavant, M. The solubility of sulphur in hydrous rhyolitic melts. *J. Petrol.* **45**, 2171–2196 (2004).
- Gaillard, F., Schmidt, B., Mackwell, S. & McCammon, C. Rate of hydrogen–iron redox exchange in silicate melts and glasses. *Geochim. Cosmochim. Acta* **67**, 2427–2441 (2003).
- Holloway, J. R. Thermodynamic modelling of geological materials: minerals, fluids and melts. *Rev. Mineral.* **17**, 211–233 (1987).
- Scaillet, B. & Pichavant, M. Role of  $f_{\text{O}_2}$  on fluid saturation in oceanic basalt. *Nature* **430**, doi:10.1038/nature02814 (published online 28 July 2004).
- Shi, P. F. & Saxena, F. K. Thermodynamic modeling of the C–H–O–S fluid system. *Am. Mineral.* **77**, 1038–1049 (1992).
- Kress, V. C. & Carmichael, I. S. E. The compressibility of silicate liquids containing  $\text{Fe}_2\text{O}_3$  and the effect of composition, temperature, oxygen fugacity and pressure on their redox states. *Contrib. Mineral. Petrol.* **108**, 82–92 (1991).
- Wallace, P. Volatiles in subduction zone magmas: concentrations and fluxes based on melt inclusion and volcanic gas data. *J. Volcanol. Geotherm. Res.* **140**, 217–240 (2004).
- Watson, E. B. Diffusion in volatile-bearing magmas. *Rev. Mineral.* **30**, 371–412 (1994).
- Westrich, H. R. & Gerlach, T. M. Magmatic gas source for the stratospheric  $\text{SO}_2$  cloud from the June 15, 1991 eruption of Mount Pinatubo. *Geology* **20**, 867–870 (1992).
- Scaillet, B. & Pichavant, M. Experimental constraints on volatile abundances in arc magmas and their implications for degassing processes. *Spec. Publ. Geol. Soc. (Lond.)* **213**, 23–52 (2003).
- Scaillet, B., Luhr, J. & Carroll, M. R. in *Volcanism and the Earth's Atmosphere* (eds Robock, A. & Oppenheimer, C.) 11–40 (Geophys. Monogr. 139, American Geophysical Union, Washington DC, 2003).
- Wallace, P., Anderson, A. T. & Davis, A. M. Quantification of pre-eruptive exsolved gas contents in silicic magmas. *Nature* **377**, 612–616 (1995).



24. Gardner, J. E., Hilton, M. & Carroll, M. R. Bubble growth in highly viscous silicate melts during continuous decompression from high pressure. *Geochim. Cosmochim. Acta* **64**, 1473–1483 (2000).
25. Eichelberger, J. C., Carrigan, C. R., Westrich, H. R. & Price, R. H. Non explosive silicic volcanism. *Nature* **323**, 598–602 (1986).
26. Symonds, R. B., Rose, W. I., Bluth, G. J. S. & Gerlach, T. M. Volcanic-gas studies: methods, results, and applications. *Rev. Mineral.* **30**, 1–66 (1994).
27. Mathez, E. A. Influence of degassing on oxidation states of basaltic magmas. *Nature* **310**, 371–375 (1984).
28. Carmichael, I. S. E. The redox states of basic and silicic magmas: a reflexion of their source regions. *Contrib. Mineral. Petrol.* **106**, 129–141 (1991).
29. Holtz, F., Behrens, H., Dingwell, D. B. & Johannes, W. H<sub>2</sub>O solubility in haplogranitic melts: compositional, pressure and temperature dependence. *Am. Mineral.* **80**, 94–108 (1995).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank M. Rutherford and P. Wallace for comments that helped us to improve our model. A.B. acknowledges support from the Swiss National Science Foundation.

**Author Contributions** A.B. incorporated the thermodynamic code of gas–melt equilibria developed by B.S. into his one-dimensional conduit flow model. A.B. performed all the simulations. Both authors contributed equally to the interpretation of the model results and to the writing of the paper.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to A.B. ([burgisse@cnr-orleans.fr](mailto:burgisse@cnr-orleans.fr)).

## LETTERS

# Evidence of giant sulphur bacteria in Neoproterozoic phosphorites

Jake V. Bailey<sup>1</sup>, Samantha B. Joye<sup>3</sup>, Karen M. Kalanetra<sup>3</sup>, Beverly E. Flood<sup>2</sup> & Frank A. Corsetti<sup>1</sup>

**In situ** phosphatization<sup>1</sup> and reductive cell division<sup>2</sup> have recently been discovered within the vacuolate sulphur-oxidizing bacteria. Here we show that certain Neoproterozoic Doushantuo Formation (about 600 million years BP) microfossils, including structures previously interpreted as the oldest known metazoan eggs and embryos<sup>3–10</sup>, can be interpreted as giant vacuolate sulphur bacteria. Sulphur bacteria of the genus *Thiomargarita* have sizes and morphologies similar to those of many Doushantuo microfossils, including symmetrical cell clusters that result from multiple stages of reductive division in three planes. We also propose that Doushantuo phosphorite precipitation was mediated by these bacteria, as shown in modern *Thiomargarita*-associated phosphogenic sites, thus providing the taphonomic conditions that preserved other fossils known from the Doushantuo Formation.

Fossil cyanobacteria<sup>11</sup>, spherical organic-walled microfossils known as acritarchs<sup>12</sup>, algae<sup>13</sup>, and putative microscopic metazoans<sup>4,7,8,14</sup> have been described from the Neoproterozoic (about 599 ± 4.2 Myr BP<sup>15</sup>) Doushantuo Formation in South China, but the unit is perhaps most noted for producing the world's oldest animal eggs (*Megasphaera*) and animal embryos (*Parapandorina*)<sup>6,9,10</sup>. Initially regarded as algae<sup>16</sup>, solitary spheroids were reinterpreted as animal eggs because of their external envelope and large size, and multiple bodies enclosed by envelopes were reinterpreted as embryos on the basis of their size and apparent evidence for reductive division by non-rigid cells<sup>9,10,17</sup>. These putative eggs and embryos, along with other Doushantuo microfossils commonly regarded as metazoans<sup>4–8,14</sup>, currently provide the earliest direct evidence of animal life.

However, several questions remain about the origins and preservational context (taphonomy) of this unprecedented accumulation of cellular microfossils. Conspicuously absent are intermediate stages linking very abundant single bodies interpreted as eggs (*Megasphaera*) and less common low-cell-number clusters interpreted as early blastula stages (*Parapandorina*), with the very rare, controversial structures interpreted as gastrulae, larvae and adult forms<sup>3,7–9,14</sup>. Furthermore, recent X-ray tomography investigations report no evidence of epithelialization by means of gastrulation in putative Doushantuo embryos, a developmental process that is ubiquitous in modern metazoan embryos<sup>18</sup>. Taphonomic studies advocate that a bias towards the preservation of early developmental stages could result from the resistance conferred by the embryonic envelope<sup>19,20</sup>. However, many specimens lack all or part of the outer envelope, a condition that if present before mineralization would result in degradation within hours<sup>20</sup>. Preservation of blastomeres within an intact fertilization envelope has been achieved in laboratory experiments, but experimental conditions required large concentrations (100 mM) of 2β-mercaptoethanol as a hypothetical simulant of extremely sulphidic conditions<sup>20</sup>. Even if conditions were capable of continuously preventing autolysis and microbial degradation of

eggs and embryos, evidence for a specific molecular mechanism of phosphatization in these cells is lacking<sup>21</sup>. Here we suggest that *Megasphaera inornata* and *Parapandorina* represent solitary and reductively dividing giant vacuolate sulphur bacteria. This explanation is consistent with the morphology, taphonomic robustness and palaeogeochemical conditions required to explain many Doushantuo globular microfossils, while providing a biogeochemical mechanism for phosphatization, which probably facilitated the preservation of other Doushantuo fossils.

In modern organic-rich marine sediments, sulphur-oxidizing bacteria, such as *Beggiatoa*, *Thioploca* and *Thiomargarita* oxidize hydrogen sulphide generated by bacterial sulphate reduction<sup>1,2,22–24</sup>. *Thiomargarita*, the focus of our investigation, is currently the largest known bacterium, with average cell diameters between 100 and 400 µm (some cells grow as large as 750 µm)<sup>22,24</sup>. *Thiomargarita* cells are generally spherical and appear hollow, with the central vacuole occupying most of the cell volume<sup>2,22,24</sup> (Figs 1a–e and 2a, b). *Thiomargarita* stores nitrate, presumably used as the oxidant for sulphide oxidation, at concentrations thousands of times that of seawater. A thin (0.5–2.0 µm thick) outer layer of cytoplasm surrounding the vacuole contains abundant spherical inclusions<sup>1,22,24</sup>. *Thiomargarita* specimens from the Gulf of Mexico occur primarily as solitary cells or clusters of cells surrounded by a mucus-filled sheath (Fig. 2a, b)<sup>2</sup>. Dense botryoidal clusters of smaller sulphur-globule-containing cells (about 20 µm) are also observed<sup>22</sup>. Whereas *Thiomargarita* cells from Namibia commonly divide in one plane, and more rarely in two planes<sup>22</sup>, recently described *Thiomargarita* cells from the Gulf of Mexico undergo division along three planes, resulting in clusters of two, four and eight cells<sup>2</sup> (Fig. 1b–e, and Supplementary Figs 1–5). No statistically significant difference was observed in the biovolumes of cell clusters, regardless of cell number, an observation consistent with reductive cellular division<sup>2</sup>. Clusters of three cells occasionally resulted from an incomplete second stage of reductive division (Fig. 1c)<sup>2</sup>.

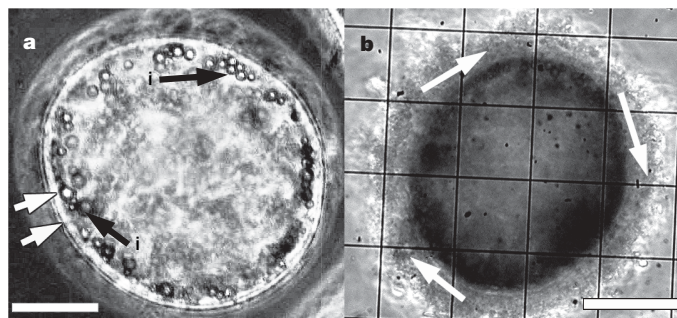
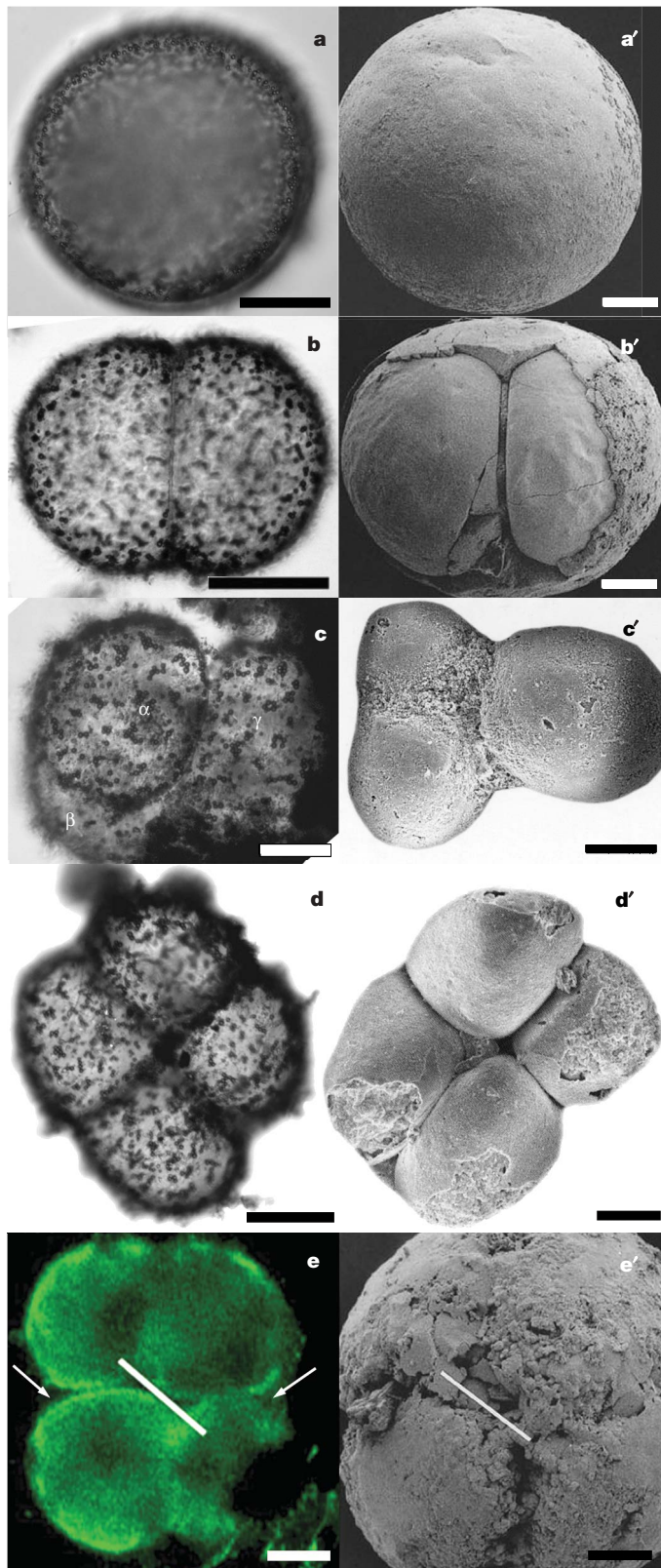
Many of the single-celled globular Doushantuo Formation microfossils exhibit sizes, morphologies and cellular division geometries consistent with those of *Thiomargarita*. The Doushantuo microfossil *Megasphaera inornata*, for example, is a large unornamented spheroidal microfossil (about 500 µm in diameter that may or may not be surrounded by a thin (about 10 µm) phosphatic envelope<sup>9</sup> (Fig. 1a'), similar to that of large single cells of *Thiomargarita* (Fig. 2a, b). In the absence of additional morphological differentiation, such spherical bodies convey limited information, highlighting a perpetual issue in the investigation of morphologically conservative fossil structures.

A more striking comparison can be made between certain *Thiomargarita* clusters and the Doushantuo microfossil known as *Parapandorina*<sup>9,10</sup>. Typically, *Parapandorina* contains even-numbered (2<sup>n</sup>) internal bodies (2, 4, 8, 16, 32) and shows little or no apparent

<sup>1</sup>Department of Earth Sciences and <sup>2</sup>Department of Biological Sciences, University of Southern California, Los Angeles, California 90089, USA. <sup>3</sup>Department of Marine Sciences, University of Georgia, Athens, Georgia 30602, USA.



change in diameter, regardless of the number of internal bodies present, suggesting that the internal bodies resulted from reductive cellular division<sup>9,10</sup> (Fig. 1b'–e'). The observation of reductive division led to their interpretation as animal embryos, as this process was at that time unknown outside the Metazoa in cells of comparable size. The recent discovery of reductive division in *Thiomargarita* now warrants a re-evaluation of the Doushantuo structures. Both *Thiomargarita* and *Parapandorina* consist of several polyhedral to



**Figure 2 | Phase contrast images of *Thiomargarita* cellular structure.** **a**, A multi-layered ultrastructure (white arrows) surrounding cytoplasm that hosts abundant spherical inclusions (labelled 'i' with black arrows) and **(b)** a mucus-filled sheath (white arrows) that hosts abundant microbial filaments. Scale bars, 100  $\mu\text{m}$  (**a**); 60  $\mu\text{m}$  (**b**).

spherical bodies surrounded by a thin envelope, show reductive cell division, and are of similar size (Fig. 1a–e, a'–e', and Supplementary Figs 1–6). *Thiomargarita* and *Parapandorina* cells both seem to be distorted by adjacent cells, suggesting a non-rigid cell wall, a fact previously used to discount the latter's interpretation as an alga<sup>17</sup> (see Supplementary Information for a detailed discussion of cell cluster geometry). *Thiomargarita* cells also include abundant inclusions and globules that can contain sulphur, polyphosphate or glycogen (labelled 'i' in Fig. 2a)<sup>22</sup> and can form larger aggregates (Supplementary Fig. 6). Subcellular structures in some *Parapandorina* specimens<sup>18</sup> may have resulted from the diagenetic alteration of such inclusions in ancient *Thiomargarita*.

Three reductive divisions (eight-cell stage) are the maximum number currently observed in *Thiomargarita*, although such division could continue to produce clusters with more cells. Reductive division in *Thiomargarita* and other bacteria is thought to be a survival response to starvation<sup>2</sup>. We postulate that *Thiomargarita* cells, if entombed within precipitating hydroxyapatite and cut off from their vital metabolic substrates, sulphide and nitrate, could continue to divide, perhaps resulting in a morphology similar to that of the rare *Parapandorina* specimens that contain more than eight cells. Of the 56 specimens examined in ref. 18, four-cell *Parapandorina* specimens were the most abundant (28%), with 74% of specimens containing eight or fewer cells. Specimens with more than 16 cells were less abundant (18%). In the 207 specimens examined in ref. 25, *Parapandorina* specimens with four internal bodies were also the most abundant (48%), and those containing more than 16 cells were absent. *Parapandorina* is much less common than *Megasphaera* in Doushantuo phosphorites, which is consistent with the interpretation of *Megasphaera inornata* as abundant solitary cells that divide reductively as a stress response, producing rare *Parapandorina*.

**Figure 1 | Comparisons of light micrograph images of translucent unmineralized modern *Thiomargarita* cells (left column) with scanning electron microscopy images of opaque mineralized Doushantuo microfossils (right column).** **a**, Solitary *Thiomargarita* cell from the Gulf of Mexico (after ref. 2). **b**, Two-cell cluster of *Thiomargarita*. **c**, Three-cell *Thiomargarita* cluster, thought to result from the incomplete division of a two-cell cluster. Greek letters identify each of the three cells. **d**, Tetragonal *Thiomargarita* tetrad resulting from reductive division in two planes. **e**, Offset between opposing cells in rhomboidal *Thiomargarita* tetrads resembles offset in some Doushantuo tetrads<sup>9</sup> and cross-furrows in four-cell blastulas. Arrows indicate thin sheath surrounding cell cluster. **a'**, *Megasphaera inornata*, from the Doushantuo Formation<sup>9</sup>. **b'**, Two appressed hemispherical bodies enclosed by an external envelope (after ref. 9). **c'**, *Thiomargarita* triplets occasionally result from incomplete division, which results in two cells with a combined volume roughly equal to the third undivided cell<sup>9</sup>. This Doushantuo globular triplet shows similar relative volumes. After ref. 5. **d'**, *Parapandorina* tetrad resulting from division in two planes. Modified after ref. 5. **e'**, Doushantuo rhomboidal tetrad modified after ref. 9. Scale bars, 150  $\mu\text{m}$  (**b'**); 100  $\mu\text{m}$  (**a**–**e**, **a'**, **c'**–**e'**).

Phosphorite deposition is relatively uncommon and poorly understood throughout the geological record. Recently, sulphide-oxidizing bacteria have been shown to mediate phosphorite deposition in modern environments<sup>1</sup>. *Thiomargarita* cell accumulations correlate with increased pore water phosphate and accumulations of phosphorus-containing minerals, which amounted to more than 50 g kg<sup>-1</sup> of dry sediment<sup>1</sup>. Laboratory experiments suggest that metabolically driven phosphate release by *Thiomargarita* controls phosphate mineral precipitation, thus providing a microbially mediated mechanism of phosphorite formation<sup>1</sup>, as was previously proposed for related sulphide-oxidizing bacteria<sup>13</sup>. Unlike animal eggs and embryos, the ambient internal and external chemical environments mediated by these microorganisms are conducive to phosphatization, providing a biogeochemical explanation for the unusually high phosphate concentrations responsible for the preservation of Doushantuo microfossils.

The unusual abundance of globular microfossils in the Doushantuo Formation has long been considered problematic for the animal embryo interpretation<sup>26</sup>; mass spawning and concentration by sedimentary processes have been proposed as possible solutions<sup>9,27</sup>. More parsimoniously, *Thiomargarita* cells of similar sizes (about 500 µm) can be found in great abundances (up to 10<sup>7</sup> cells m<sup>-2</sup>) and would not require unusual circumstances to explain large accumulations of their fossils<sup>24</sup> (see Supplementary Information). In addition, stages of reductive division in *Thiomargarita* are separated by months to years<sup>2</sup>, allowing a longer window for the observed preferential preservation of two-cell, four-cell and eight-cell clusters, unlike invertebrate embryos that develop from the fertilized egg, through the blastula stages to a gastrula in a matter of a few hours<sup>28</sup>. The extremely long intervals between reductive division in *Thiomargarita* might also explain why cell clusters with more than eight cells have not yet been observed.

Figure 3a shows *Thiomargarita* cells from Namibia covered by filamentous bacteria interpreted to be symbiotic sulphate-reducing bacteria<sup>22</sup>. *Thiomargarita* cells from the Gulf of Mexico also commonly occur with filamentous and spheroidal bacteria. Doushantuo globular microfossils, but not acritarchs and algae, are commonly associated with phosphatized filaments and spheroids previously interpreted to be fossilized infesting bacteria<sup>21</sup> (Fig. 3b). Taphonomic studies indicate that infesting bacteria rapidly destroy embryonic animal cells<sup>20</sup>, making the preferential co-preservation of embryos and the offending degradative bacteria unlikely, although not unprecedented<sup>21</sup>. Conversely, the preservation of bacteria with their natural host *Thiomargarita* cells requires no such fortuitous mineralization of transient putrefying remains.

The geochemical conditions during the deposition of the Doushantuo Formation, but not necessarily the palaeoenvironment itself, may have been analogous to modern upwelling zones where *Thiomargarita* was discovered<sup>24</sup>. Multiple lines of palaeontological

and lithological evidence point to enriched nutrient concentrations in Doushantuo depositional environments<sup>12</sup>. Additionally, indicators of sulphidic conditions, which would have been a requirement for *Thiomargarita*, are abundant in the Doushantuo Formation. Sulphur and oxygen isotope values of sulphate from Doushantuo Formation phosphorites provide evidence of probable bacterial sulphate reduction, perhaps coupled to sulphide oxidation<sup>29</sup>. Abundant pyrite in phosphorites and in microfossils<sup>21</sup> is also suggestive of sulphidic conditions. In an environment with nutrient-rich, productive surface waters and euxinic bottom waters, the remains of planktonic algae, and perhaps metazoans, from the overlying water column, along with allochthonous benthic organic debris, would occur together with abundant sulphide-oxidizing bacteria as in the modern counterparts, resulting in fossil assemblages rich in phosphatic spheroidal and globular forms, such as those observed in the Doushantuo Formation.

We do not suggest that the Doushantuo microbiota is composed entirely of sulphur bacterial remains, because there are many structures in the Doushantuo that clearly do not resemble sulphur bacteria. Furthermore, the hypothesis posed here does not invalidate the possibility that some Doushantuo globular microfossils are indeed animal eggs and embryos, because the two hypotheses are not mutually exclusive. Instead, our reassessment provides a morphologically plausible alternative interpretation of the most abundant globular microfossils (*Megasphaera inornata* and *Parapandorina*) as giant vacuolate sulphur bacteria, and offers a mechanism of preservation by means of microbially mediated phosphate release by *Thiomargarita*, as noted in modern habitats. Given the possibility that *Megasphaera inornata* and *Parapandorina* represent giant bacteria, and the controversy surrounding other putative metazoans in the Doushantuo Formation<sup>3,30</sup>, the interpretation of Doushantuo globular microfossils as metazoan embryos deserves further scrutiny.

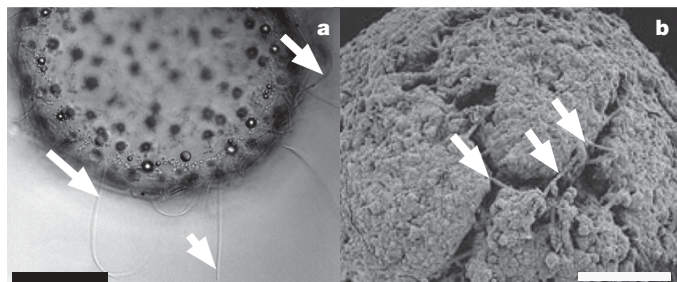
## METHODS

For *Thiomargarita* sample collection, sediment samples from a variety of hydrocarbon seeps in the Gulf of Mexico were collected during a research expedition on the R/V *Seward Johnson* in July 2002. Push cores were collected with the *Johnson Sea-Link II* submersible<sup>2</sup>. Sediment samples were preserved in 2.5% glutaraldehyde at 4 °C until microscopic analysis. Microscopic images and measurements of cell diameters were made with phase-contrast microscopy, with the exception of Fig. 1e, which was stained with fluorescein isothiocyanate (FITC) and imaged with a Leica TCS-SP Laser Scanning Confocal Microscope fitted with an argon ion laser (488 nm).

Received 26 July; accepted 20 November 2006.

Published online 20 December 2006.

- Schulz, H. N. & Schulz, H. D. Large sulfur bacteria and the formation of phosphorite. *Science* **307**, 416–418 (2005).
- Kalanetra, K. M., Joye, S. B., Sunseri, N. R. & Nelson, D. C. Novel vacuolate sulfur bacteria from the Gulf of Mexico reproduce by reductive division in three dimensions. *Environ. Microbiol.* **7**, 1451–1460 (2005).
- Xiao, S., Yuan, X. & Knoll, A. H. Eumetazoan fossils in terminal Proterozoic phosphorites? *Proc. Natl Acad. Sci. USA* **97**, 13684–13689 (2000).
- Li, C., Chen, J. & Hua, T. E. Precambrian sponges with cellular structures. *Science* **279**, 879–882 (1998).
- Chen, J.-Y. *The Dawn of Animal World* (Jiangsu Science & Technology, Nanjing, 2004).
- Chen, J.-Y. *et al.* Phosphatized polar lobe-forming embryos from the Precambrian of southwest China. *Science* **312**, 1644–1646 (2006).
- Chen, J.-Y. *et al.* Precambrian animal life: probable developmental and adult cnidarian forms from southwest China. *Dev. Biol.* **248**, 182–196 (2002).
- Chen, J.-Y. *et al.* Precambrian animal diversity: New evidence from high resolution phosphatized embryos. *Proc. Natl Acad. Sci. USA* **97**, 4457–4462 (2000).
- Xiao, S. & Knoll, A. H. Phosphatized animal embryos from the Neoproterozoic Doushantuo Formation at Weng'an, Guizhou, South China. *J. Paleontol.* **74**, 767–788 (2000).
- Xiao, S., Zhang, Y. & Knoll, A. H. Three-dimensional preservation of algae and animal embryos in a Neoproterozoic phosphorite. *Nature* **391**, 553–558 (1998).
- Awramik, S. M. *et al.* Prokaryotic and eukaryotic microfossils from a Proterozoic/Phanerozoic transition in China. *Nature* **315**, 655–658 (1985).
- Zhou, C., Brasier, M. D. & Xue, Y. Three-dimensional phosphatic preservation of giant acritarchs from the Terminal Proterozoic Doushantuo Formation in Guizhou and Hubei Provinces, South China. *Palaeontology* **44**, 1157–1178 (2001).



**Figure 3 | Filaments associated with *Thiomargarita* cells and Doushantuo microfossils.** **a**, *Thiomargarita* samples from Namibia (modified after ref. 22) and the Gulf of Mexico are often covered with filamentous bacteria that are probably sulphate-reducing symbionts. **b**, Doushantuo globular microfossils, such as this *Parapandorina* specimen, are commonly covered with filaments thought to be mineralized embryo-infesting bacteria (after ref. 9). Scale bars, 50 µm (**a**); 100 µm (**b**).



13. Zhang, Y., Yin, L., Xiao, S. & Knoll, A. H. Permineralized fossils from the Terminal Proterozoic Doushantuo Formation, South China. *Paleontol. Soc. Mem.* **50**, 1–52 (1998).
14. Chen, J.-Y. *et al.* Small bilaterian fossils from 40 to 55 million years before the Cambrian. *Science* **305**, 218–222 (2004).
15. Barfod, G. H. *et al.* New Lu-Hf and Pb-Pb age constraints on the earliest animal fossils. *Earth Planet. Sci. Lett.* **201**, 203–212 (2002).
16. Xue, Y., Tang, T., Yu, C. & Zhou, C. Large spheroidal chlorophyta fossils from the Doushantuo Formation phosphoric sequence (late Sinian), central Guizhou, South China. *Acta Palaeontol. Sin.* **34**, 688–706 (1995).
17. Xiao, S. Mitotic topologies and mechanics of Neoproterozoic algae and animal embryos. *Paleobiology* **28**, 244–250 (2002).
18. Hagadorn, J. W. *et al.* Cellular and subcellular structure of Neoproterozoic animal embryos. *Science* **314**, 291–294 (2006).
19. Martin, D., Briggs, D. E. G. & Parkes, R. J. Decay and mineralization of invertebrate eggs. *Palaos* **20**, 562–572 (2005).
20. Raff, E. *et al.* Experimental taphonomy shows the feasibility of fossil embryos. *Proc. Natl Acad. Sci. USA* **103**, 5846–5851 (2006).
21. Xiao, S. & Knoll, A. H. Fossil preservation in the Neoproterozoic Doushantuo phosphorite Lagerstätte, South China. *Lethaia* **32**, 219–240 (1999).
22. Schulz, H. N. in *The Prokaryotes: An Evolving Electronic Resource for the Microbiological Community* (eds Dworkin, M. *et al.*) (<http://141.150.157.117:8080/prokPUB/chaprender/jsp/showchap.jsp?chapnum=513>) (2006).
23. Jørgensen, B. B. & Nelson, D. C. in *Sulfur Biogeochemistry—Past and Present* (eds Amend, J. P., Edwards, K. J. & Lyons, T. W.) 63–81 (Geol. Soc. Am. Spec. Pap. 379, Boulder, Colorado, 2004).
24. Schulz, H. N. *et al.* Dense populations of a giant sulfur bacterium in Namibian shelf sediments. *Science* **284**, 493–495 (1999).
25. Dornbos, S. Q. *et al.* Precambrian animal life: Taphonomy of phosphatized metazoan embryos from southwest China. *Lethaia* **38**, 101–109 (2005).
26. Xue, Y. S., Tang, T. F. & Yu, C. L. 'Animal embryos,' a misinterpretation of Neoproterozoic microfossils. *Acta Micropaleontol. Sin.* **16**, 1–4 (1999).
27. Xiao, S. & Knoll, A. H. Embryos or algae? A reply. *Acta Micropaleontol. Sin.* **16**, 313–323 (1999).
28. Costello, D. P. & Henley, C. *Methods for Obtaining and Handling Marine Eggs and Embryos* (Marine Biological Laboratory, Woods Hole, Massachusetts, 1971).
29. Goldberg, T., Poulten, S. W. & Strauss, H. Sulphur and oxygen isotope signatures of late Neoproterozoic to early Cambrian sulphate, Yangtze Platform, China: Diagenetic constraints and seawater evolution. *Precamb. Res.* **137**, 223–241 (2005).
30. Bengston, S. & Budd, G. Comment on 'Small bilaterian fossils from 40 to 55 million years before the Cambrian'. *Science* **306**, 1291a (2004).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank D. Bottjer, D. Caron, A. Jones, R. Schaffner, S. Douglas, A. Thompson and D. Nelson for discussions, advice and/or assistance. This project was funded by the US National Science Foundation Graduate Research Fellowship, Earth Sciences, and Life in Extreme Environments programs, as well as the NASA Exobiology Program. Submersible operations in the Gulf of Mexico were supported by the US Department of Energy and the National Oceanic and Atmospheric Administration's National Undersea Research Program.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.V.B. ([jvbailey@usc.edu](mailto:jvbailey@usc.edu)).

## LETTERS

# Habitat modification alters the structure of tropical host–parasitoid food webs

Jason M. Tylianakis<sup>1†</sup>, Teja Tscharntke<sup>1</sup> & Owen T. Lewis<sup>2</sup>

Global conversion of natural habitats to agriculture has led to marked changes in species diversity and composition<sup>1</sup>. However, it is less clear how habitat modification affects interactions among species<sup>2</sup>. Networks of feeding interactions (food webs) describe the underlying structure of ecological communities, and might be crucially linked to their stability and function<sup>3–7</sup>. Here, we analyse 48 quantitative food webs<sup>8,9</sup> for cavity-nesting bees, wasps and their parasitoids across five tropical habitat types. We found marked changes in food-web structure<sup>10,11</sup> across the modification gradient, despite little variation in species richness. The evenness of interaction frequencies declined with habitat modification, with most energy flowing along one or a few pathways in intensively managed agricultural habitats. In modified habitats there was a higher ratio of parasitoid to host species and increased parasitism rates, with implications for the important ecosystem services, such as pollination and biological control, that are performed by host bees and wasps<sup>12</sup>. The most abundant parasitoid species was more specialized in modified habitats, with reduced attack rates on alternative hosts. Conventional community descriptors failed to discriminate adequately among habitats, indicating that perturbation of the structure and function of ecological communities might be overlooked in studies that do not document and quantify species interactions. Altered interaction structure therefore represents an insidious and functionally important hidden effect of habitat modification by humans.

Habitat modification is the leading global cause of species extinctions and alterations in abundance of persisting species, and is a particular threat to biodiversity in extremely diverse tropical ecosystems<sup>13</sup>. All species are linked in networks of mutualistic and antagonistic interactions<sup>14</sup>. The importance of conserving these interactions and associated processes, as well as the component species, has been stressed repeatedly<sup>2,15,16</sup>, particularly as humans rely on ecosystem services associated with species interactions, such as pollination and biological control<sup>17</sup>. Research on ‘interaction extinctions’ and altered patterns of interactions in ecological communities has been impeded by the lack of appropriate analytical tools, and by a failure to quantify the frequency or even to detect the presence of interspecific interactions<sup>18</sup>. Early food-web studies treated all interactions as equal, irrespective of their magnitude or frequency, and qualitative properties used to describe food webs were found to be highly sensitive to variation in sampling effort<sup>19,20</sup>.

Recently, methods have been developed to allow the quantification of interactions at the community scale<sup>9</sup>, thereby giving a more robust description of community structure, and providing insights into the dynamic processes that structure ecological communities<sup>21</sup>. Although this approach is most commonly applied to networks of trophic interactions (quantitative food webs<sup>9,22</sup>), it is equally applicable to other interspecific interactions such as pollination<sup>23</sup>. Existing

studies have used quantitative food webs to describe the structures of relatively intact species assemblages, but differences in quantitative food-web structure across habitats remain unknown. Nevertheless, these methods have considerable potential for quantifying the effect of human activities on networks of interacting species<sup>24</sup>.

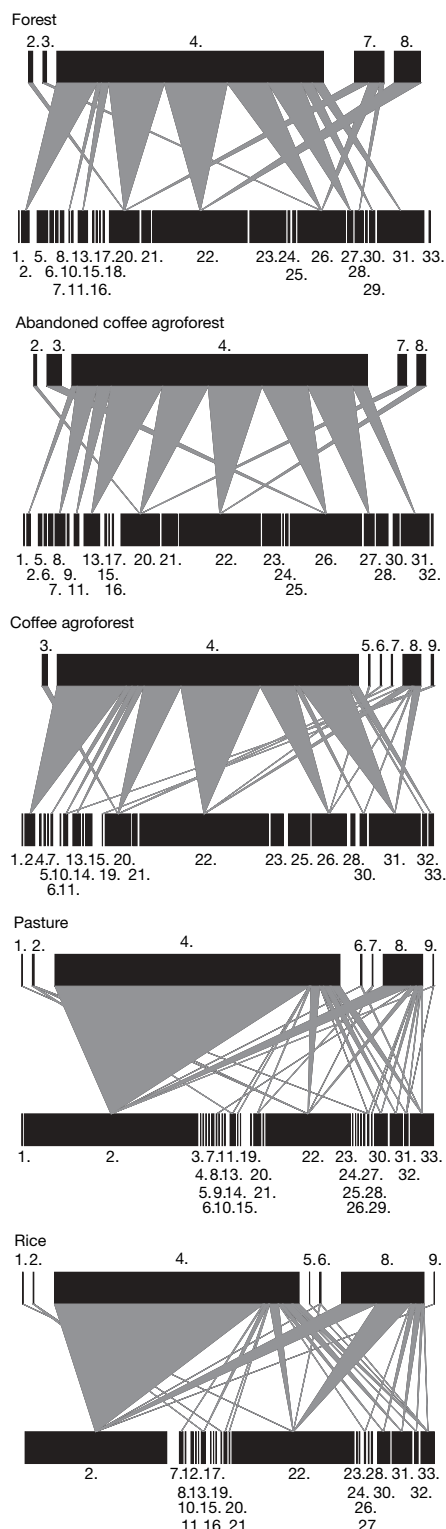
The information contained in food webs can be summarized in various ways. Quantitative, weighted equivalents of the food-web statistics that were developed to describe binary webs have now been derived<sup>10</sup>. Here, we use four of these quantitative metrics, with well known qualitative counterparts (connectance, linkage density ( $LD_q$ ), generality ( $G_q$ ) and vulnerability ( $V_q$ )), and quantitative measures of interaction evenness and compartmentalization (see Supplementary Methods 1), to assess the effect of a habitat modification gradient on food-web structure in the tropical biodiversity hotspot of coastal Ecuador. Our food webs document 4,090 trophic interactions involving 33 species of bees and wasps (Hymenoptera: Apidae, Megachilidae, Mutilidae, Pompilidae, Sphecidae and Vespidae) and nine parasitoid and kleptoparasite (collectively referred to hereafter as parasitoid) species (Hymenoptera: Eulophidae, Ichneumonidae, Leucospidae, Megachilidae and Chrysididae; Diptera: Bombyliidae).

The bees and wasps in our focal food webs perform important ecosystem services<sup>12</sup>. Bees are key pollinators of both crops and wild plants, and even small variations in the number of species can have severe effects on pollination rates<sup>25</sup>. Functional differences between species allow complementarity in pollination, such that extinction or altered abundance of key species can have disproportionate effects<sup>25</sup>. Similarly, predatory wasps can be important predators of herbivorous insects, such as pest caterpillars<sup>12</sup>. The effects of changes in land use on the diversity of bees and wasps have been well documented<sup>25–27</sup>. However, little is known about the structure of interactions between these species and their natural enemies, and how this might affect associated ecosystem services.

We found striking (Fig. 1) and highly significant differences in quantitative food-web structure among habitats (multivariate general linear model (GLM): Wilks’ lambda = 0.128,  $F_{20,124} = 5.30$ ,  $P < 0.0001$ ). Food-web quantitative vulnerability (a measure of the ratio of host to parasitoid species) and interaction evenness (a measure of the uniformity of energy flows along different pathways; see Supplementary Methods 1 for methods of calculation for these and other metrics) showed significant differences among habitats (Table 1). Importantly, differences in food web metrics did not result simply from differences in species richness across habitats: habitat explained a statistically significant portion of variation in the metrics even after controlling for the effects of host and parasitoid richness (multivariate GLM: Wilks’ lambda for effect of habitat = 0.34,  $F_{20,124} = 2.399$ ,  $P = 0.002$ , Table 1); and host and parasitoid species richness per site did not vary significantly across habitats (ANOVA:  $F_{4,43} < 2.5$ ,  $P > 0.05$  in both cases). Rather, habitat modification led

<sup>1</sup>Agroecology, Georg August University, Waldweg 26, Goettingen D-37073, Germany. <sup>2</sup>Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK. <sup>†</sup>Present address: School of Biological Sciences, University of Canterbury, Private Bag 4800, Christchurch 8020, New Zealand.





**Figure 1 | Quantitative host-parasitoid food webs along a gradient of increasing habitat modification (top to bottom).** For each web, lower bars represent host (bee and wasp) abundance and upper bars represent parasitoid abundance, drawn at different scales. Linkage width indicates frequency of each trophic interaction. As a summary, the webs show interaction data pooled across all replicates for each habitat type, although analyses were conducted on a per-site basis. Species codes are given in Supplementary Table 3. Host and parasitoid order is consistent across webs, and the webs are drawn at different scales (for total host abundances, see Supplementary Fig. 1). *Melittobia acasta* is parasitoid species 4; *Anthidium* sp. is host species 2.

**Table 1 | Effects of habitat type on parasitism rate and individual food web metrics**

Variable	$F_{4,41}$	Model 1 $P$	$r^2$	$F_{4,41}$	Model 2 $P$	$r^2$
Parasitism rate*†	15.54	<0.00001	0.579	13.24	<0.00001	0.329
Linkage density	1.32	0.278	0.103	1.24	0.308	0.107
Generality	2.11	0.096	0.158	2.01	0.110	0.131
Vulnerability†	10.08	<0.00001	0.478	9.85	0.00001	0.326
Evenness†	5.85	<0.0001	0.357	5.97	<0.001	0.223
Connectance	2.04	0.106	0.129	1.59	0.195	0.084
Compartment diversity‡	0.48	0.749	0.045	0.49	0.741	0.046

Results derived from independent GLMs. Model 1 has habitat type entering the model first, before parasitoid and host diversity. Model 2 has habitat type entering last, after removal of variation explained by parasitoid and host diversity.

\* Arcsine square root transformed.

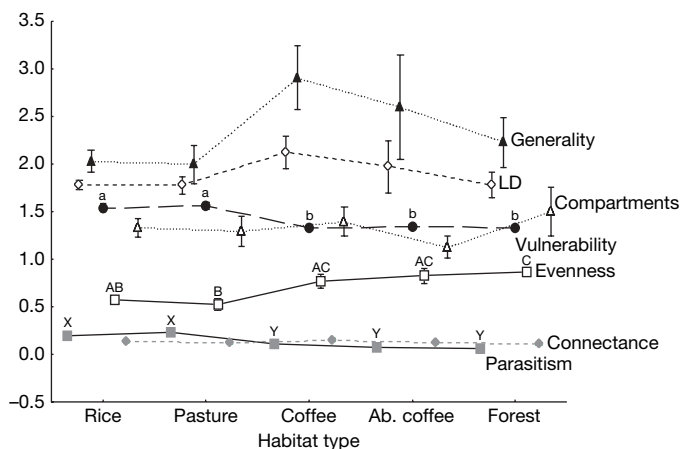
† Significant at a Bonferroni-corrected  $\alpha$  of 0.0071.

‡ Log<sub>10</sub> transformed.

to large differences in the evenness of interactions (Table 1, Fig. 2). Quantitative weighted vulnerability increased greatly in more modified systems (Table 1, Fig. 2) as did the proportion of individual host bees and wasps that were parasitized (Table 1, Fig. 2). We found no significant difference in linkage density or compartment diversity across habitats (Table 1).

For completeness, we repeated these analyses with qualitative metrics based on binary presence/absence interaction data (see Supplementary Table 1, Supplementary Fig. 1). These metrics showed significant variation among habitats, but failed to reveal the differences observed in the quantitative webs. Rather, they were strongly biased by sample size<sup>11,19</sup>, and showed significantly higher linkage density, generality, vulnerability, interaction richness and connectance in rice and pasture, because the abundances of both hosts and parasitoids was two to five times higher in these habitats than in the less modified systems. This contrasts sharply with the higher vulnerability and interaction evenness in less modified systems, using quantitative, weighted measures that are not affected by sample size differences.

Inspection of the quantitative food webs (Fig. 1) indicates the specific changes that are responsible for differences in food-web structure among habitats. Food webs in intensive agricultural systems (rice and pasture) were dominated by a single trophic interaction (a gregarious pupal parasitoid *Melittobia acasta* Walk. (Chalcidoidea: Eulophidae) parasitising a bee *Anthidium* sp. (Hymenoptera: Anthidiinae); parasitoid species 4 and host species 2 in Fig. 1). This shift towards reduced interaction evenness would be



**Figure 2 | The effects of habitat modification on rates of parasitism (mean  $\pm$  s.e.m.) and six quantitative food web metrics (see Supplementary Methods 1): linkage density (LD), generality, vulnerability, evenness, connectance and compartment diversity.** Letters above individual means indicate significant differences among habitat types for that particular metric. Letters in common or no letters indicates no significant difference.

impossible to detect with qualitative metrics (see Supplementary Table 1, Supplementary Fig. 1), despite its conspicuousness in the quantitative webs (Fig. 1). *Anthidium* sp. is characteristic of intensive agricultural systems, with 98% of the 7,839 observed individuals found in rice or pasture. *Melittobia acasta* attacked 16 host species across the different habitats, but *Anthidium* sp. was the host for 78% of *M. acasta* individuals, and following removal of *Anthidium* sp. from the data set, vulnerability and evenness showed no significant variation across habitats (see Supplementary Table 2).

However, the high abundance of *Anthidium* sp. in modified habitats was not the only factor driving the changes in food-web structure. Rather, *M. acasta*, the most abundant parasitoid in all habitats, was more specialized in modified habitats, with the evenness (see Supplementary Methods 2) of host species used being significantly lower in rice and pasture (ANOVA:  $F_{4,41} = 7.54$ ,  $P = 0.0001$ ,  $r^2 = 0.424$ ). This specialization on *Anthidium* sp. means that alternative hosts might benefit. Indeed, *Megachile* sp. 1, the host species with the second highest frequency of parasitism from *M. acasta* (11.6% of individuals, compared to 32.6% in *Anthidium* sp.), experiences proportionately reduced parasitism in plots where parasitism of *Anthidium* sp. is high, even after controlling for differences in parasitism among habitats (see Supplementary Methods 3, multivariate GLM, effect of *Anthidium* sp. parasitism rates:  $F_{1,9} = 16.19$ ,  $P = 0.003$ ,  $r^2 = 0.484$ ). *Anthidium* sp. in disturbed habitats might act as an abundant source of the highly generalist *M. acasta*, with possible consequences for other host species in agricultural or adjacent natural habitats through spillover of parasitoids<sup>28</sup>. The low specificity of *M. acasta* could allow parasitoid-mediated apparent competition to occur between host species, as has been documented in another tropical host–parasitoid food web<sup>21</sup>.

Both quantitative vulnerability (the weighted ratio of parasitoid to host species) and levels of parasitism were significantly higher in rice and pasture habitats. High parasitoid diversity can promote high rates of parasitism<sup>29</sup>, so high vulnerability can be expected to correlate with increased parasitism rates. Furthermore, parasitism was highest in the habitats with lower interaction evenness, consistent with data from the biological control literature showing the strongest top–down control in modified habitats, where food webs were dominated by a single link<sup>30</sup>. As solitary bees are known to be important pollinators of both wild and crop plants, and wasps provide biological control by consuming pest species such as caterpillars<sup>12,25</sup>, increased mortality of these guilds in heavily modified systems might be detrimental to the maintenance of the ecosystem services they provide.

No significant differences in either quantitative linkage density or food web compartmentalization were detected among habitats. Linkage density is strongly affected by species richness<sup>19</sup>, and the absence of significant differences in richness among habitats might have led to the consistency of this metric. Compartmentalization was consistently low across habitats, largely because there was little tendency for individual parasitoid species to specialize on individual host species and because hosts were typically vulnerable to multiple parasitoid species in each habitat.

Although the most severely modified habitats, pasture and rice, showed characteristic differences in food-web structure relative to other habitats (Fig. 1), both managed and abandoned coffee agroforests had similar food-web structures to forest, with no significant differences observed for any of the food-web metrics (Fig. 2). This is consistent with previous studies that have shown the biodiversity of agroforests to be comparable with that of natural forests<sup>26,27</sup>, and indicates that a threshold level of habitat modification or openness, caused by removal of tree cover, might cause a significant step-change in species diversity<sup>26,27</sup> and composition (see Supplementary Fig. 2), with an associated shift in food-web structure. Coffee provides an intermediate habitat, harbouring interactions that are typical of both heavily modified and unmodified systems. Thus, appropriately managed agroecosystems can retain ecological

assemblages that are structurally and functionally similar to unmodified habitats. By contrast, highly modified pasture and rice habitats show marked changes in quantitative food-web structure, even where there has been little overall effect on the diversity of component species. Our results indicate that studies of human influences such as habitat change, fragmentation and climate change that focus solely on diversity and richness metrics might overlook important alterations to community structure, species interactions and ecosystem functions.

## METHODS

**Study region.** The 48 study plots were spread across three cantons in the region of Jipijapa, Manabí province, Southwest Ecuador. Individual plot location, age and area details are given in ref. 26. A map of the region is given in ref. 27. This area falls within the Choco-Manabí region (a biodiversity hotspot), but large-scale agricultural conversion threatens the local biodiversity and the ecosystem services it provides<sup>26,29</sup>.

We compared the five predominant land use types in the region. In order of decreasing modification, these were (with percentage of study zone occupied): rice (1.9% including other arable crops), pasture (16.2%), coffee agroforest (54.8%), abandoned coffee agroforests and forests (17.2% combined; not distinguishable using satellite imagery)<sup>27</sup>. Coffee agroforests can vary greatly with individual management and for consistency we selected plots that had been in cultivation for more than 8 years, where no chemical fertilizers had been used, and which had similar shade tree cover (60–80%) comprising a mixture of remnant forest species, legumes (for example, *Inga* spp.) planted for nitrogen enrichment, and various trees planted for fruit or timber. The understorey herb layer in these habitats is cleared twice per year. Sampling was carried out in twelve replicate sites of each of the managed systems, six forest and six abandoned coffee agroforest sites. Each site had experienced the current management regime for at least five years, and on average, there was no significant difference in the area of sites from each land use type (exact age and area of the sites can be found in the electronic supplement to ref. 26).

**Trap nests.** Nine trap nests, each comprising a 22-cm long, 15-cm diameter plastic tube with reed internodes inserted, were positioned in each of the 48 plots, to provide standardized nesting sites for naturally occurring bee, wasp and natural enemy assemblages<sup>26</sup>. Trap nests were evaluated every month from June 2003 to October 2004. Data from each of the nine traps per plot were pooled across all months for analyses. Species lists of primary trap-nesting species (hosts) and parasitoids are presented in Supplementary Table 2. As both solitary (one parasitoid individual per host individual) and gregarious (multiple parasitoid individuals per host individual) parasitoids were reared, we defined interaction frequency as the number of host larvae that were parasitized. As the dominant parasitoid is gregarious, an alternative food-web representation based on the number of parasitoid individuals recorded leads to a more markedly uneven distribution of interaction frequencies in modified habitats.

**Analyses.** Analyses were conducted in Statistica 6.1 (Statsoft). Several of the food-web metrics were intercorrelated (see Supplementary Table 4), so, to reduce the probability of a type I error, we tested the effect of habitat on food-web metrics ( $LD_q$ ,  $G_q$ ,  $V_q$ , connectance, compartment diversity and interaction evenness) per site using a general linear model (GLM) with multivariate responses, then made subsequent individual analyses of each response variable and rates of parasitism. To account for any effects on food web structure of among-habitat variation in species richness of hosts and parasitoids, we fitted GLMs with habitat type as a fixed factor and parasitoid and host richness as covariates. We used a Bonferroni corrected  $\alpha$  of 0.0071. By introducing habitat type into the model before or after host and parasitoid diversity, using type I sums of squares, it was possible to quantify the overall variance in food web metrics across habitat types, as well as the variance after controlling for differences in host and parasitoid diversity among habitats. Results after controlling for parasitoid and host abundance were qualitatively the same, but for completeness we present results from both models for each metric (Table 1). We made *post hoc* comparisons using a Tukey Honestly Significant Difference test for unequal sample sizes. Error degrees of freedom for multivariate tests were calculated using denominator synthesis, and values presented are rounded to the nearest integer. Model residuals were tested for adherence to a normal distribution and homogeneity of variances. Compartment diversity was  $\log_{10}$  transformed, and proportion parasitism was arcsine square root transformed to meet assumptions of the tests. Residuals from all models were also tested for spatial autocorrelation using a Mantel test conducted in R (R Development Core Team; <http://www.R-project.org>), with a distance matrix constructed from  $x$  and  $y$  global positioning system coordinates. The tests used Pearson correlations and a Bonferroni corrected  $\alpha$  of 0.0071. No



significant autocorrelation was found (Mantel statistic  $r < 0.023$ ,  $P > 0.01$  in all cases).

Received 14 October; accepted 3 November 2006.

1. Tilman, D. *et al.* Forecasting agriculturally driven global environmental change. *Science* **292**, 281–284 (2001).
2. van der Putten, W. H. *et al.* Trophic interactions in a changing world. *Basic Appl. Ecol.* **5**, 487–494 (2004).
3. McCann, K. The diversity–stability debate. *Nature* **405**, 228–233 (2000).
4. Melián, C. J. & Bascompte, J. Food web structure and habitat loss. *Ecol. Lett.* **5**, 37–46 (2002).
5. Ives, A. R. & Cardinal, B. J. Food-web interactions govern the resistance of communities after non-random extinctions. *Nature* **429**, 174–177 (2004).
6. Rooney, N., McCann, K., Gellner, G. & Moore, J. C. Structural asymmetry and the stability of diverse food webs. *Nature* **442**, 265–269 (2006).
7. Montoya, J. M., Pimm, S. L. & Solé, R. V. Ecological networks and their fragility. *Nature* **442**, 259–264 (2006).
8. Memmott, J., Godfray, H. C. J. & Gauld, I. D. The structure of a tropical host–parasitoid community. *J. Anim. Ecol.* **63**, 521–540 (1994).
9. Müller, C. B., Adriaanse, I. C. T., Belshaw, R. & Godfray, H. C. J. The structure of an aphid–parasitoid community. *J. Anim. Ecol.* **68**, 346–370 (1999).
10. Bersier, L. F., Banasek-Richter, C. & Cattin, M. F. Quantitative descriptors of food-web matrices. *Ecology* **83**, 2394–2407 (2002).
11. Banasek-Richter, C., Cattin, M. F. & Bersier, L. F. Sampling effects and the robustness of quantitative and qualitative food-web descriptors. *J. Theor. Biol.* **226**, 23–32 (2004).
12. Losey, J. E. & Vaughan, M. The economic value of ecological services provided by insects. *Bioscience* **56**, 311–323 (2006).
13. Foley, J. A. *et al.* Global consequences of land use. *Science* **309**, 570–574 (2005).
14. Paine, R. T. Food webs: Road maps of interactions or grist for theoretical development? *Ecology* **69**, 1648–1654 (1988).
15. Janzen, D. H. The deflowering of Central America. *Nat. Hist.* **83**, 48–53 (1974).
16. Kearns, C. A., Inouye, D. W. & Waser, N. M. Endangered mutualisms: the conservation of plant–pollinator interactions. *Annu. Rev. Ecol. Syst.* **29**, 83–112 (1998).
17. Costanza, R. *et al.* The value of the world's ecosystem services and natural capital. *Nature* **387**, 253–260 (1997).
18. Cohen, J. E. *et al.* Improving food webs. *Ecology* **74**, 252–258 (1993).
19. Goldwasser, L. & Roughgarden, J. Sampling effects and the estimation of food-web properties. *Ecology* **78**, 41–54 (1997).
20. Martinez, N. D., Hawkins, B. A., Dawah, H. A. & Feifarek, B. P. Effects of sampling effort on characterization of food-web structure. *Ecology* **80**, 1044–1055 (1999).
21. Morris, R. J., Lewis, O. T. & Godfray, H. C. J. Experimental evidence for apparent competition in a tropical forest food web. *Nature* **428**, 310–313 (2004).
22. Lewis, O. T. *et al.* Structure of a diverse tropical forest insect–parasitoid community. *J. Anim. Ecol.* **71**, 855–873 (2002).
23. Memmott, J. The structure of a plant–pollinator food web. *Ecol. Lett.* **2**, 276–280 (1999).
24. Henneman, M. L. & Memmott, J. Infiltration of a Hawaiian community by introduced biological control agents. *Science* **293**, 1314–1316 (2001).
25. Kremen, C., Williams, N. M. & Thorp, R. W. Crop pollination from native bees at risk from agricultural intensification. *Proc. Natl Acad. Sci. USA* **99**, 16812–16816 (2002).
26. Tylianakis, J. M., Klein, A. M. & Tscharntke, T. Spatiotemporal variation in the effects of a tropical habitat gradient on Hymenoptera diversity. *Ecology* **86**, 3296–3302 (2005).
27. Tylianakis, J. M., Klein, A. M. & Tscharntke, T. Spatial scale of observation affects  $\alpha$ ,  $\beta$  and  $\gamma$  diversity of cavity-nesting bees and wasps across a tropical land use gradient. *J. Biogeogr.* **33**, 1295–1304 (2006).
28. Rand, T., Tylianakis, J. M. & Tscharntke, T. Spillover edge effects: the dispersal of agriculturally-subsidized insect natural enemies into adjacent natural habitats. *Ecol. Lett.* **9**, 603–614 (2006).
29. Tylianakis, J. M., Tscharntke, T. & Klein, A. M. Diversity, ecosystem function and stability of parasitoid–host interactions across a tropical habitat gradient. *Ecology* **87**, 3047–3057 (2006).
30. Hawkins, B. A., Mills, N. J., Jervis, M. A. & Price, P. W. Is the biological control of insects a natural phenomenon? *Oikos* **86**, 493–506 (1999).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank J. Bascompte, M. B. Bonsall, M. M. Bos, H. C. J. Godfray, B. A. Hawkins, R. J. Morris, C. B. Müller and T. Roslin for discussions or comments, and J. Casquete, J. Pico, G. Sacoto, C. Valarezo, C. Calderon, A. Choez, J. Lino, R. Olschewski, G. H. J. de Koning and B. Pico for assistance with field or laboratory work and coordination. H. C. J. Godfray provided software and advice for drawing food webs and calculating compartmentalization. This research was funded by the Federal Ministry for Research and Education (BMBF) Germany, and a Royal Society University Research Fellowship to O.T.L.

**Author Contributions** J.M.T. was responsible for writing, field data collection, research planning and analysis; T.T. for research planning, site selection and funding application; and O.T.L. for writing, analysis and preparation of food webs (Fig. 1). All authors discussed the results and edited the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.M.T. ([jason.tylianakis@canterbury.ac.nz](mailto:jason.tylianakis@canterbury.ac.nz)).

## LETTERS

# A stomatin-domain protein essential for touch sensation in the mouse

Christiane Wetzel<sup>1</sup>, Jing Hu<sup>1\*</sup>, Dieter Riethmacher<sup>2\*</sup>, Anne Benckendorff<sup>1\*</sup>, Lena Harder<sup>1</sup>, Andreas Eilers<sup>1</sup>, Rabih Moshourab<sup>1</sup>, Alexey Kozlenkov<sup>1</sup>, Dominika Labuz<sup>3</sup>, Ombretta Caspani<sup>3</sup>, Bettina Erdmann<sup>4</sup>, Halina Machelska<sup>3</sup>, Paul A. Heppenstall<sup>1,3</sup> & Gary R. Lewin<sup>1</sup>

Touch and mechanical pain are first detected at our largest sensory surface, the skin. The cell bodies of sensory neurons that detect such stimuli are located in the dorsal root ganglia, and subtypes of these neurons are specialized to detect specific modalities of mechanical stimuli. Molecules have been identified that are necessary for mechanosensation in invertebrates but so far not in mammals. In *Caenorhabditis elegans*, *mec-2* is one of several genes identified in a screen for touch insensitivity and encodes an integral membrane protein with a stomatin homology domain<sup>1</sup>. Here we show that about 35% of skin mechanoreceptors do not respond to mechanical stimuli in mice with a mutation in stomatin-like protein 3 (SLP3, also called Stoml3), a mammalian *mec-2* homologue that is expressed in sensory neurons. In addition, mechanosensitive ion channels found in many sensory neurons do not function without SLP3. Tactile-driven behaviours are also impaired in *SLP3* mutant mice, including touch-evoked pain caused by neuropathic injury. SLP3 is therefore indispensable for the function of a subset of cutaneous mechanoreceptors, and our data support the idea that this protein is an essential subunit of a mammalian mechanotransducer.

Cutaneous sensory neurons detect mechanical stimuli in the periphery through specialized mechanotransduction ion channels that may form a part of a larger complex of proteins<sup>2,3</sup>. Studies in *C. elegans* have shown that the *mec-2* gene, which encodes a stomatin-domain-containing protein, is required for touch receptor function<sup>1,4</sup>. The MEC-2 protein is postulated to participate in a complex that includes the Deg/ENaC channels, MEC-4 and MEC-10, as well as the MEC-6 protein<sup>2,5,6</sup>. Stomatin (band 7.2b), first isolated from human red blood cells<sup>7,8</sup>, is the prototypical stomatin-domain protein<sup>9,10</sup> expressed in mouse dorsal root ganglion (DRG) neurons<sup>9</sup>. Electrophysiological recordings in stomatin mutant mice indicate a mild impairment but not a complete loss of sensory neuron mechanosensitivity (C. Martinez-Salgado, A.B., C.W. and G.R.L., unpublished observations). We identified the complementary DNA of SLP3 (ref. 11), an integral membrane protein, as being expressed by DRG neurons. SLP3 and stomatin have similar degrees of sequence identity to MEC-2 (Supplementary Fig. 1). *In situ* hybridization and quantitative reverse transcriptase-mediated polymerase chain reaction (RT-PCR) showed *SLP3* messenger RNA to be expressed by all DRG neurons and to be selectively expressed in neuronal tissues (Supplementary Fig. 2).

We generated *SLP3* mutant mice (Supplementary Fig. 3) that were viable and fertile. To test mechanoreceptor function we used an *in vitro* skin nerve preparation to record from single cutaneous sensory neurons in the saphenous nerve (single-unit recording)<sup>12</sup>. Normally more than 90% of sensory fibres innervating the skin are mechano-

receptors. Mechano-insensitive fibres can be identified only with an electrical search technique (Fig. 1a)<sup>13</sup> (see Methods). With this technique the conduction velocity of single units was measured, permitting their classification as A $\beta$  fibres, A $\delta$  fibres or unmyelinated C fibres. For each single unit an extensive search was made for its mechanosensitive receptive field. In contrast with wild-type littermates, 30 and 40% of A $\delta$  and A $\beta$  fibres, respectively, lacked all mechanosensitivity in *SLP3*<sup>-/-</sup> mice (Fig. 1a). Among C fibres we observed no change in mechanoreceptive properties (Fig. 1a, d). Sensory axons without mechanosensitivity appeared otherwise normal (Supplementary Fig. 4a).

A $\beta$  and A $\delta$  fibres can be classified into four major mechanoreceptor types<sup>14</sup>, three of which are low-threshold mechanoreceptors (Supplementary Table 1). Low-threshold mechanoreceptors include slowly adapting mechanoreceptors (SAMs) and rapidly adapting mechanoreceptors (RAMs) and D-hair mechanoreceptors. The fourth group, high-threshold mechanoreceptors, are often referred to as A $\delta$ -fibre mechanonociceptors. We went on to characterize the remaining functional mechanoreceptors in the *SLP3* mutant (Fig. 1b, and Supplementary Table 1). If any one of the above receptors selectively lost mechanosensitivity, a shift in the proportion of RAMs and SAMs in the A $\beta$  range or A $\delta$ -fibre mechanonociceptors, and D hairs in the A $\delta$  range, would be expected. In fact the relative incidence of these mechanoreceptors was unchanged (Fig. 1b, Supplementary Table 1). Thus, SLP3 seems to be required for the function of 30–40% of the myelinated mechanoreceptors regardless of receptor type. It is still possible that mechanoreceptors affected by the *SLP3* mutation represent undetected subpopulations of receptors. For example, in rodents some RAMs are ‘field’ receptors not driven by hair movement<sup>14,15</sup>, and these cannot be distinguished from other RAMs in the *in vitro* skin nerve preparation.

In *SLP3*<sup>-/-</sup> mice a large proportion (about 45%) of putative RAM fibres were unresponsive to even very fast movements (1.2–2.9  $\mu\text{m ms}^{-1}$ ). These units responded only to brisk tapping of the receptive field with a blunt probe and were designated ‘tap’ units. Less than 5% of RAM fibres were tap units in control mice (Fig. 1c). The conduction velocities of the more frequent tap fibres (about 40%) encountered in *SLP3*<sup>-/-</sup> mice were no different from those of normal mechanosensitive A $\beta$  fibres (tap conduction velocity  $15.8 \pm 0.8 \text{ m s}^{-1}$ , compared with  $15.7 \pm 1.0 \text{ m s}^{-1}$  for normal RAM fibres in the *SLP3*<sup>-/-</sup> mice; see Supplementary Table 1). The stimulus response properties of SAM fibres in mutant mice were also mildly but significantly impaired (Supplementary Fig. 4c–e). The remaining ‘non-tap’ RAMs and all D-hair receptors recorded in *SLP3*<sup>-/-</sup> mice

<sup>1</sup>Department of Neuroscience, Max-Delbrück Center for Molecular Medicine and Charité Universitätsmedizin Berlin, Robert-Rössle-Strasse 10, 13125 Berlin-Buch, Germany. <sup>2</sup>Zentrum für Molekulare Neurobiologie, Universität Hamburg, Falkenried 94, 20251 Hamburg, Germany. <sup>3</sup>Klinik für Anaesthesiologie und Operative Intensivmedizin, Charité Universitätsmedizin Berlin, Campus Benjamin Franklin, Hindenburgdamm 30, 12200 Berlin, Germany. <sup>4</sup>Electronmicroscopy, Max-Delbrück Center for Molecular Medicine, Robert-Rössle-Strasse 10, 13125 Berlin-Buch, Germany.

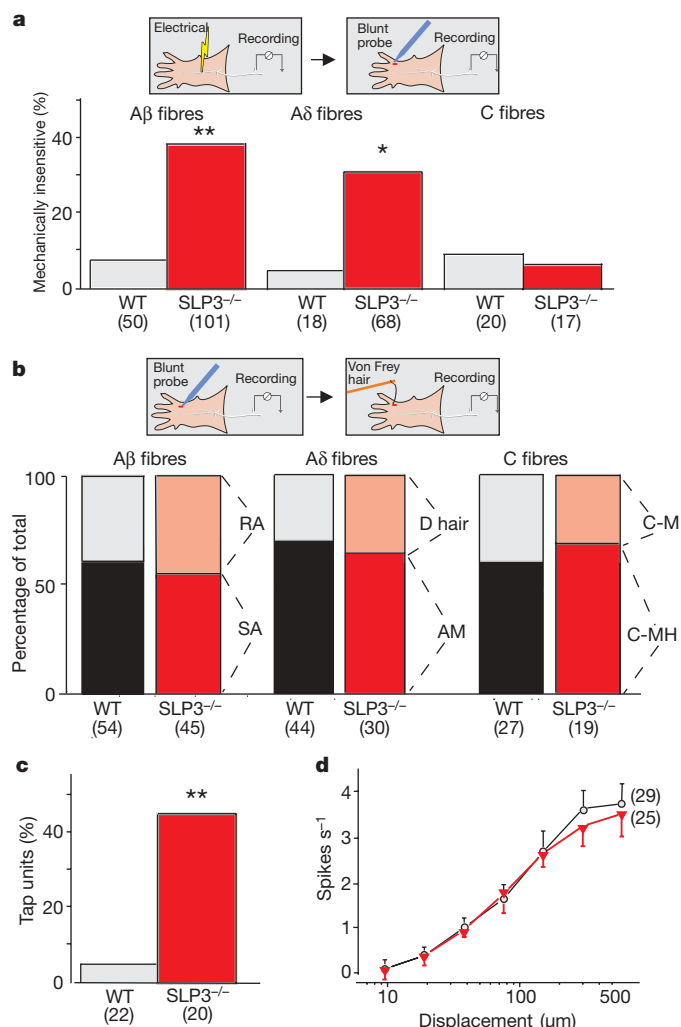
\*These authors contributed equally to this work.



had normal mechanosensitivity (Supplementary Fig. 4). A $\delta$ -fibre mechanonociceptors and C-fibre nociceptors both displayed normal mechanosensitivity in *SLP3* mutant mice (Fig. 1d, and data not shown).

We tested whether the lack of mechanosensitivity in *SLP3* mutants is caused by a loss of sensory neurons, by a withdrawal of axons from the peripheral nerve or by changes in skin innervation. An anatomical analysis using light and electron microscopy revealed no loss of sensory fibres or their terminal endings in the skin of *SLP3* mutant mice (Supplementary Fig. 5).

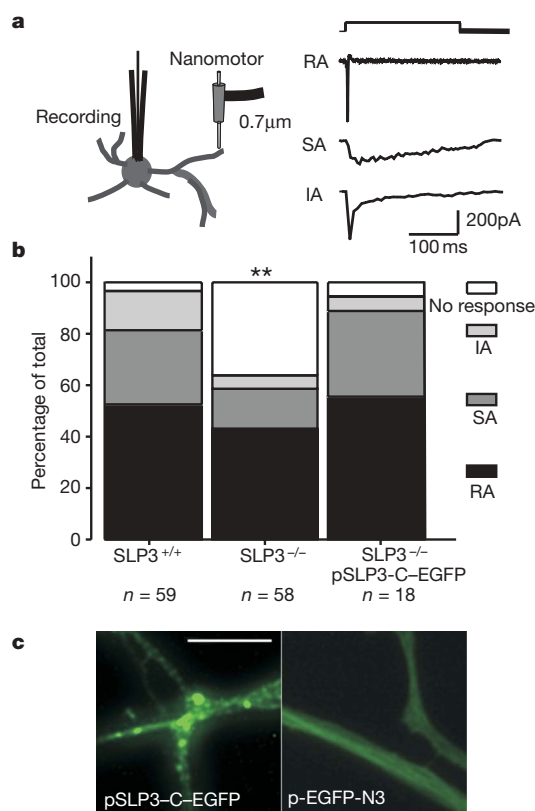
We next tested whether *SLP3*, like *C. elegans* MEC-2 (ref. 4), is required for the generation of mechanosensitive currents. The neurites and cell soma of acutely isolated mouse sensory neurons possess mechanosensitive channels<sup>16–18</sup>, and neuritic currents are activated by displacements of less than 750 nm with latencies of about 600  $\mu$ s (ref. 16). With the use of such stimuli, more than 95% of isolated sensory neurons possessed one of three mechanically activated currents,



**Figure 1 | Mechanically insensitive primary afferents.** **a**, The electrical search protocol is illustrated in the insets. Note the increase in mechanically insensitive A fibres in *SLP3*<sup>-/-</sup> mice (asterisk,  $P < 0.05$  (A $\delta$ ); two asterisks,  $P < 0.005$  (A $\beta$ );  $\chi^2$  test), but no change in mechanosensitive C fibres;  $n = 18$ –31 mice per genotype. WT, wild-type. **b**, Using the illustrated mechanical search protocol no significant difference ( $P > 0.05$ ;  $\chi^2$  test) in the proportions of mechanoreceptor types in A $\beta$ , A $\delta$  (AM) and C fibres was found;  $n = 8$ –10 mice per genotype. RA, rapidly adapting; SA, slowly adapting; C-M, C-mechanoreceptor; C-MH, C-mechanohair nociceptor. **c**, Increase in the proportion of RAM fibres in *SLP3*<sup>-/-</sup> with a tap response (two asterisks,  $P < 0.005$ ;  $\chi^2$  test). **d**, Normal stimulus response functions of C-fibre nociceptors in *SLP3*<sup>-/-</sup> mice. Black circles, *SLP3*<sup>+/+</sup>; red triangles, *SLP3*<sup>-/-</sup>. The number of neurons recorded is indicated in parentheses in each panel. Error bars indicate s.e.m.

rapidly adapting (inactivation in less than 2 ms), intermediate-adapting (inactivation in less than 50 ms) and slowly adapting (no adaptation during a 230-ms stimulus) (Fig. 2a). About 36% of sensory neurons from *SLP3*<sup>-/-</sup> mice had no current response to mechanical stimulation of their neurites, in contrast with less than 5% in control cultures (Fig. 2b). The proportion of cells that displayed slowly adapting or rapidly adapting mechanically activated currents was slightly reduced, indicating that *SLP3* might be required for both types of current. The amplitude and kinetics of the mechanosensitive current in the remaining *SLP3*<sup>-/-</sup> neurons were indistinguishable from those of control neurons (Supplementary Table 2). Next we transfected *SLP3* mutant neurons with a mouse *SLP3* cDNA tagged at the carboxy terminus with green fluorescent protein (*SLP3*-C-EGFP). Fluorescently labelled *SLP3* was observed in a punctate pattern in the newly grown neurites starting at 15 h in culture, and almost all labelled neurons had a mechanosensitive current (17 out of 18 cells; Fig. 2b, c, and Supplementary Table 2). Control EGFP constructs (Fig. 2c) had no effect on the incidence of mechanosensitive currents (four out of nine labelled cells had no mechanosensitive current).

We observed no change in the properties of voltage-gated channels in *SLP3* mutant neurons (data not shown). However, *SLP3* mutant neurons had a slightly more depolarized resting membrane potential than wild-type neurons ( $-56.6 \pm 1.5$  mV in *SLP3*<sup>-/-</sup> neurons in comparison with  $-62.0 \pm 1.0$  mV in *SLP3*<sup>+/+</sup> neurons;  $P < 0.005$ , Student's *t*-test). However, the resting membrane potential of



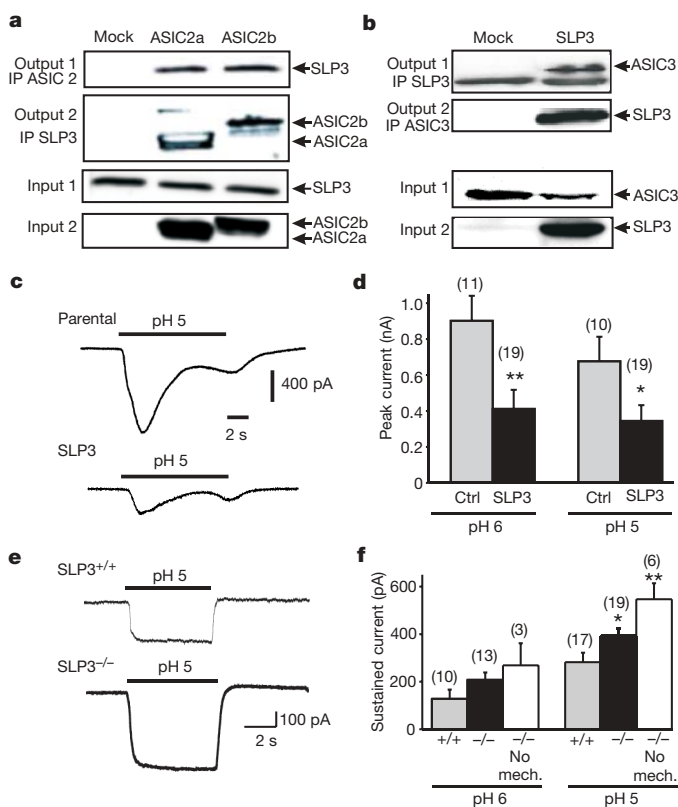
**Figure 2 | *SLP3* is required for the expression of mechanosensitive currents in subsets of DRG neurons.** **a**, Three types of mechanically activated current, designated rapidly adapting (RA), slowly adapting (SA) or intermediate-adapting (IA), observed by using displacement stimuli (750 nm) of the neurite. **b**, Stacked histogram showing proportions of sensory neurons with the different mechanically activated currents from *SLP3*<sup>+/+</sup>, *SLP3*<sup>-/-</sup> and *SLP3*<sup>-/-</sup> mice transfected with an EGFP-tagged *SLP3* cDNA. Note a significant increase in the proportion of neurons lacking a mechanosensitive current in *SLP3*<sup>-/-</sup> neurons (two asterisks,  $P < 0.005$ ;  $\chi^2$  test). Mechanically activated currents were restored to normal proportions in cells expressing *SLP3*-C-EGFP protein. The numbers of cells tested are indicated. **c**, Left: cells transfected with pSLP3-C-EGFP showed labelled *SLP3* in puncta throughout the neurites. Right: expression of EGFP in sensory neurites. Scale bar, 5  $\mu$ m.

$SLP3^{-/-}$  neurons with or without mechanosensitive current was not different (Supplementary Table 2).

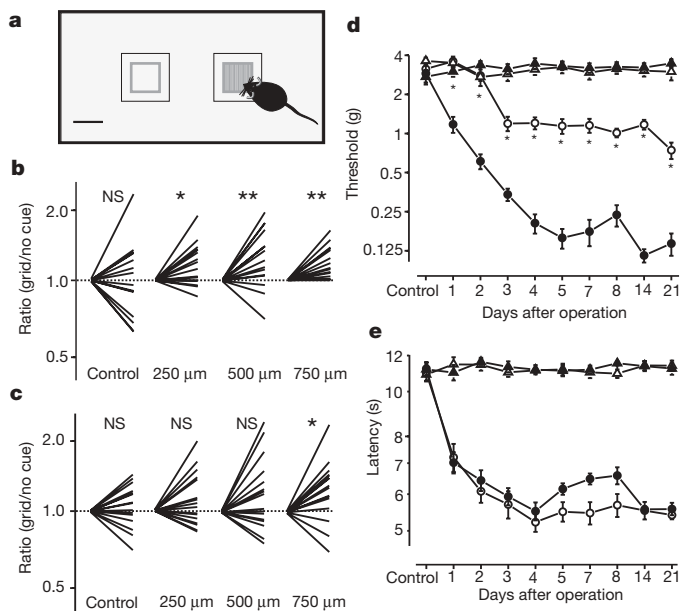
Mutations in genes that encode the acid-sensitive ion channels ASIC2a, ASIC2b and ASIC3 alter mechanoreceptor sensitivity *in vivo*<sup>19–22</sup>. We therefore tested whether the SLP3 protein can also interact with members of the ASIC family. We showed that ASIC2a, ASIC2b and ASIC3 are immunoprecipitated with SLP3 when overexpressed in HEK-293 cells (Fig. 3a, b). TRP-V1, a thermosensitive member of the *trp* channel family<sup>23</sup>, did not immunoprecipitate with SLP3 (data not shown). HEK-293 cells possess amiloride-sensitive proton-gated currents with properties similar to those described for heteromeric ASIC channels<sup>24</sup> (Fig. 3c, and Supplementary Fig. 6). When rat SLP3 was expressed in HEK-293 cells, there was a clear inhibition of proton gating of the endogenously expressed ASICs that was similar to that reported for stomatin<sup>25</sup> (Fig. 3c, d). We also found that sustained proton-gated currents were larger in SLP3 mutant sensory neurons (Fig. 3e). The proton-gated currents in the large and medium-sized cells tested were sensitive to benzamil, an agent that blocks ASICs, and the cells were insensitive to

capsaicin (data not shown). The increased amplitude of proton-gated currents in SLP3<sup>-/-</sup> neurons was observed in cells with and without mechanosensitive currents (Fig. 3f). Thus, SLP3 can physically associate and modulate channels that have been implicated in mechanotransduction.

Taken together, these results indicated that SLP3 mutant mice might have reduced tactile acuity, but no suitable behavioural tests were available to test this idea in mice. We therefore developed new behavioural tests to examine this possibility (Fig. 4a, and Supplementary Fig. 7). We tested the ability of mice to detect and respond to a grating cue placed on the floor of an arena in complete darkness (see Methods for details). Control C57BL/6N mice showed significant interest in all three grids tested, as reflected by a positive ratio of distance travelled in the area of the cue compared with no cue (Fig. 4b). SLP3<sup>-/-</sup> mice (C57BL/6N genetic background) did not show any significant tendency ( $P > 0.05$ , paired *t*-test) to prefer the finest grids used in the task (250 and 500  $\mu$ m), but SLP3<sup>-/-</sup> mice did respond significantly to the coarsest grid used (750  $\mu$ m) (Fig. 4b, c). Thus, the loss of mechanoreceptor function in SLP3<sup>-/-</sup> mice correlates with a reduction in tactile discrimination capability. SLP3<sup>-/-</sup> mice performed largely normally in a motor task (the accelerating rotorod test), which indicates that proprioception is not severely affected in these mice (Supplementary Fig. 8).



**Figure 3 | Association of SLP3 with ASIC ion channels.** **a**, Interaction of SLP3 with both ASIC2a and ASIC2b. HEK-293 cells expressing Myc-tagged SLP3 (input 1) were transfected with control (mock), Flag-tagged ASIC2a or Flag-tagged ASIC2b (input 2). Immunoprecipitates containing ASIC2a or ASIC2b were blotted and probed to detect SLP3 (output 1). Both ASIC2a and ASIC2b were detected in lysates immunoprecipitated with anti-Myc (output 2). IP, immunoprecipitation. **b**, HEK-293 cells expressing a Flag-tagged ASIC3 (input 1) were transfected with control (mock) or Myc-tagged SLP3 (input 2). Immunoprecipitates containing ASIC3 and SLP3 were detected in western blot (output 1). ASIC3 is detected in immunoprecipitates containing SLP3 (output 2). **c**, Proton-gated inward currents shown in parental HEK-293 cells and HEK-293 cells stably expressing SLP3. **d**, Mean amplitude of proton-gated current in control and SLP3-expressing HEK-293 cells. **e**, Proton-gated currents in large DRG neurons from SLP3<sup>+/+</sup> and SLP3<sup>-/-</sup> mice. Amplitude of sustained current at pH 5 was clearly larger in SLP3<sup>-/-</sup> neurons than in wild-type neurons. **f**, All neurons, with a mechanosensitive current or without one (no mech.), had larger proton-gated currents at pH 5 and pH 6. Asterisk,  $P < 0.05$ ; two asterisks,  $P < 0.01$ ; Mann–Whitney *U*-test. Error bars indicate s.e.m.



**Figure 4 | Tactile driven behaviour is altered in SLP3 mutant mice.**

**a**, Schematic drawing of the experimental arrangement, showing how the grid cue was placed in one of two spatially equivalent positions. Scale bar, 4 cm. **b**, The ratio of the distance travelled between no cue and grid cue (no cue set to 1.0) is shown as a single line for each of 16 SLP3<sup>+/+</sup> mice (more than 1.0 indicates more time on the grid; less than 1.0 indicates less time). Control mice spent more time in the vicinity of the cue for all grids tested (asterisk,  $P < 0.05$ ; two asterisks,  $P < 0.005$ ; paired *t*-test). **c**, SLP3<sup>-/-</sup> mice exhibited a significant preference only for the coarsest grid (750  $\mu$ m). Tests with no cue indicated no preference for either location (control). **d**, Mechanical thresholds (von Frey test) for foot withdrawal plotted before and after the induction of CCI (circles) or sham surgery (triangles). SLP3<sup>-/-</sup> mice (open symbols) developed significantly less mechanical allodynia than controls throughout the entire time course after CCI (asterisk,  $P < 0.05$ ; two-way repeated-measures ANOVA and Bonferroni *t*-test). Further, SLP3<sup>-/-</sup> mice developed mechanical allodynia later than SLP3<sup>+/+</sup> mice (filled symbols) ( $P < 0.05$ ; one-way repeated-measures ANOVA and Dunnett's test). There were no significant changes in sham-operated mice of either genotype ( $P > 0.05$ ; two-way repeated-measures ANOVA). **e**, No significant differences in thermal hyperalgesia between SLP3<sup>-/-</sup> and SLP3<sup>+/+</sup> mice (two-way repeated-measures ANOVA;  $P > 0.05$ ). The symbols used are the same as those in **d**. Error bars in **d** and **e** indicate s.e.m.

Stimulation of low-threshold mechanoreceptors by brushing the skin can, under neuropathic conditions (after nerve injury), lead to intense pain in humans and animals<sup>14,26</sup>. We subjected control C57BL/6N mice and *SLP3*<sup>-/-</sup> mice (C57BL/6N background) to a unilateral chronic constriction injury of the sciatic nerve (CCI)<sup>26</sup>. Within the first few days, control mice developed a full-blown allodynia with mechanical thresholds for paw withdrawal decreasing from 3 g to less than 0.25 g. In contrast, *SLP3*<sup>-/-</sup> mice did not show any hyperalgesia until the third day after the lesion, and mechanical thresholds decreased to just above 1 g until the end of the observation period (21 days) (Fig. 4d). No impairment of the CCI-induced thermal hyperalgesia was observed in *SLP3*<sup>-/-</sup> mice (Fig. 4e). Thus, targeting essential mechanotransduction molecules such as SLP3 might provide a novel peripheral means of neuropathic pain control.

We have shown that SLP3 is absolutely required for the function of many mechanoreceptors in the mouse. In contrast, polymodal nociceptors in the DRG are not significantly affected by the mutation of *SLP3* (Figs 1 and 2). Thus, many but not all sensory neurons require SLP3 for normal mechanoreceptor function and to maintain mechanosensitive channels (Fig. 2, and Supplementary Table 2). Our data are therefore consistent with a model in which about one-third of sensory neurons transduce mechanical stimuli by using mechanosensitive ion channels that depend on SLP3 for their function. SLP3 can physically interact with and modulate the gating of ASIC channels that have previously been implicated in the transduction of mechanical stimuli by sensory neurons<sup>19,22,27</sup>. Our data therefore indicate that SLP3 might act as an essential subunit in a complex that includes these channels or other, as yet unidentified, mechanosensitive channels.

## METHODS

See Supplementary Information for detailed methods.

**Molecular biology and electrophysiology.** The rat *SLP3* cDNA was cloned from mRNA isolated from DRG with the use of similar methods to those described previously<sup>9</sup>. The *SLP3* mutant mouse was generated with standard gene-targeting methods<sup>28</sup>. Single-unit recordings with a skin nerve preparation and patch-clamp experiments with cultured DRG neurons were performed in accordance with previously published procedures<sup>16,29</sup>.

**Tactile acuity test.** The MoTil system (TSE Systems) was used to monitor open-field behaviour of mice in complete darkness. After an acclimatization period of 10 min, the target area (Fig. 4a) was replaced with a tactile cue (grid) and the animals' behaviour was monitored for the next 90 min. The distance travelled in the vicinity of the grid and no cue area was expressed as a ratio with the control area set to a value of 1.0. More time spent in the vicinity of the grid is reflected as a ratio more than 1.0, and less time as less than 1.0. Control experiments in which no cue was placed in either area indicated that the time spent in either area was randomly distributed between ratios more than 1.0 and less than 1.0 (Fig. 4b). Each mouse was tested with one tactile cue, once per day, and the order was randomized. Equal numbers of male and female mice were used.

**Sensory neuron transfection.** Cells were transiently transfected with the use of the commercially available Nucleofector system (Amaxa Biosystems). In brief, neurons were suspended in 100 µl of Mouse Neuron Nucleofector Solution and 5 µg of plasmid DNA at 22–25 °C. The mixture was transferred to a cuvette for electroporation with program C-13. After electroporation the cell suspension was transferred to 500 µl of RPMI 1640 medium (Gibco) for 10 min at 37 °C. This suspension, supplemented with 10% horse serum, was used to plate the cells onto glass coverslips for recording. The RPMI medium was replaced with the standard DRG medium 3–4 h later.

Received 25 July; accepted 30 October 2006.

Published online 13 December 2006.

- Huang, M., Gu, G., Ferguson, E. L. & Chalfie, M. A stomatin-like protein necessary for mechanosensation in *C. elegans*. *Nature* **378**, 292–295 (1995).
- Ernstrom, G. G. & Chalfie, M. Genetics of sensory mechanotransduction. *Annu. Rev. Genet.* **36**, 411–453 (2002).
- Gillespie, P. G. & Walker, R. G. Molecular basis of mechanosensory transduction. *Nature* **413**, 194–202 (2001).
- O'Hagan, R., Chalfie, M. & Goodman, M. B. The MEC-4 DEG/ENaC channel of *Caenorhabditis elegans* touch receptor neurons transduces mechanical signals. *Nature Neurosci.* **8**, 43–50 (2005).
- Goodman, M. B. et al. MEC-2 regulates *C. elegans* DEG/ENaC channels needed for mechanosensation. *Nature* **415**, 1039–1042 (2002).

- Chelur, D. S. et al. The mechanosensory protein MEC-6 is a subunit of the *C. elegans* touch-cell degenerin channel. *Nature* **420**, 669–673 (2002).
- Stewart, G. W. et al. Isolation of cDNA coding for an ubiquitous membrane protein deficient in high Na<sup>+</sup>, low K<sup>+</sup> stomatocytic erythrocytes. *Blood* **79**, 1593–1601 (1992).
- Hiebl-Dirschmied, C. M. et al. Cloning and nucleotide sequence of cDNA encoding human erythrocyte band 7 integral membrane protein. *Biochim. Biophys. Acta* **1090**, 123–124 (1991).
- Mannsfeldt, A. G., Carroll, P., Stucky, C. L. & Lewin, G. R. Stomatin, a MEC-2 like protein, is expressed by mammalian sensory neurons. *Mol. Cell. Neurosci.* **13**, 391–404 (1999).
- Fricke, B. et al. Stomatin immunoreactivity in ciliated cells of the human airway epithelium. *Anat. Embryol. (Berl.)* **207**, 1–7 (2003).
- Goldstein, B. J., Kulaga, H. M. & Reed, R. R. Cloning and characterization of SLP3: a novel member of the stomatin family expressed by olfactory receptor neurons. *J. Assoc. Res. Otolaryngol.* **4**, 74–82 (2003).
- Koltzenburg, M., Stucky, C. L. & Lewin, G. R. Receptive properties of mouse sensory neurons innervating hairy skin. *J. Neurophysiol.* **78**, 1841–1850 (1997).
- Kress, M., Koltzenburg, M., Reeh, P. W. & Handwerker, H. O. Responsiveness and functional attributes of electrically localized terminals of cutaneous C-fibers in vivo and in vitro. *J. Neurophysiol.* **68**, 581–595 (1992).
- Lewin, G. R. & Moshourab, R. Mechanosensation and pain. *J. Neurobiol.* **61**, 30–44 (2004).
- Lewin, G. R. & McMahon, S. B. Physiological properties of primary sensory neurons appropriately and inappropriately innervating skin in the adult rat. *J. Neurophysiol.* **66**, 1205–1217 (1991).
- Hu, J. & Lewin, G. R. Mechanosensitive currents in the neurites of cultured mouse sensory neurones. *J. Physiol. (Lond.)* advance online publication, 12 October 2006 (doi:10.1113/jphysiol.2006.117648).
- McCarter, G. C., Reichling, D. B. & Levine, J. D. Mechanical transduction by rat dorsal root ganglion neurons in vitro. *Neurosci. Lett.* **273**, 179–182 (1999).
- Drew, L. J. et al. Acid-sensing ion channels ASIC2 and ASIC3 do not contribute to mechanically activated currents in mammalian sensory neurones. *J. Physiol. (Lond.)* **556**, 691–710 (2004).
- Price, M. P. et al. The mammalian sodium channel BNC1 is required for normal touch sensation. *Nature* **407**, 1007–1011 (2000).
- Alvarez de la Rosa, D., Zhang, P., Shao, D., White, F. & Canessa, C. M. Functional implications of the localization and activity of acid-sensitive channels in rat peripheral nervous system. *Proc. Natl Acad. Sci. USA* **99**, 2326–2331 (2002).
- Garcia-Anoveros, J., Samad, T. A., Zúvela-Jelaska, L., Woolf, C. J. & Corey, D. P. Transport and localization of the DEG/ENaC ion channel BNaC1alpha to peripheral mechanosensory terminals of dorsal root ganglia neurons. *J. Neurosci.* **21**, 2678–2686 (2001).
- Price, M. P. et al. The DRASIC cation channel contributes to the detection of cutaneous touch and acid stimuli in mice. *Neuron* **32**, 1071–1083 (2001).
- Caterina, M. J. & Julius, D. The vanilloid receptor: a molecular gateway to the pain pathway. *Annu. Rev. Neurosci.* **24**, 487–517 (2001).
- Gunthorpe, M. J., Smith, G. D., Davis, J. B. & Randall, A. D. Characterisation of a human acid-sensing ion channel (hASIC1a) endogenously expressed in HEK293 cells. *Pflugers Arch.* **442**, 668–674 (2001).
- Price, M. P., Thompson, R. J., Eshcol, J. O., Wemmie, J. A. & Benson, C. J. Stomatin modulates gating of acid-sensing ion channels. *J. Biol. Chem.* **279**, 53886–53891 (2004).
- Tal, M. & Bennett, G. J. Extra-territorial pain in rats with a peripheral mononeuropathy: mechano-hyperalgesia and mechano-allodynia in the territory of an uninjured nerve. *Pain* **57**, 375–382 (1994).
- Page, A. J. et al. The ion channel ASIC1 contributes to visceral but not cutaneous mechanoreceptor function. *Gastroenterology* **127**, 1739–1747 (2004).
- Brockschneider, D. et al. Cell depletion due to diphtheria toxin fragment A after Cre-mediated recombination. *Mol. Cell. Biol.* **24**, 7636–7642 (2004).
- Shin, J. B., Martinez-Salgado, C., Heppenstall, P. A. & Lewin, G. R. A T-type calcium channel required for normal function of a mammalian mechanoreceptor. *Nature Neurosci.* **6**, 724–730 (2003).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank A. Scheer, H. Thränhardt and K. Borgwald for technical support; G. Kempermann for advice on confocal microscopy and the rotorod test; and C. Birchmeier, A. Hammes, I. Ibanez-Tallon and T. Willnow for helpful comments on the manuscript. The Deutsche Forschungsgemeinschaft provided grant support. J.H. was supported by a von Humboldt fellowship.

**Author Contributions** C.W., J.H. and G.R.L. performed electrophysiological experiments on *SLP3* mutants. D.R. and A.B. generated the *SLP3* mutant. C.W. characterized the *SLP3* mouse. L.H. and C.W. performed the behavioural experiments. A.E. and R.M. performed biochemical interaction studies, and O.C. and P.H. characterized *SLP3* effects in HEK-293 cells. A.B. cloned the rat *SLP3* cDNA. J.H. and A.K. conducted rescue experiments. B.E. performed electron microscopy. H.M. and D.L. conducted neuropathic pain measurements in the *SLP3* mutant mice. G.R.L. conceived and planned experimental studies with help from C.W. and P.A.H. G.R.L. wrote the paper.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to G.R.L. ([glewin@mdc-berlin.de](mailto:glewin@mdc-berlin.de)).



## LETTERS

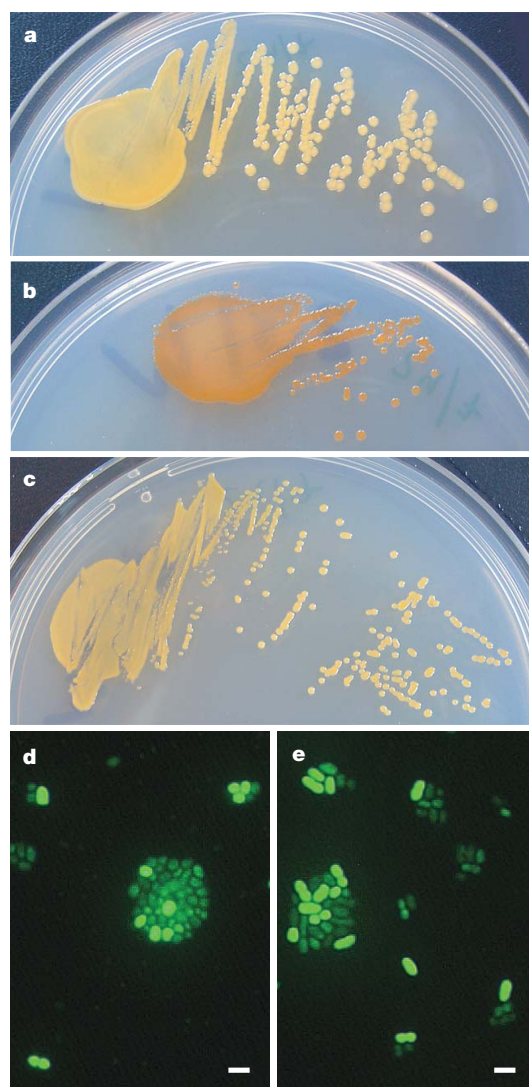
# Light stimulates growth of proteorhodopsin-containing marine Flavobacteria

Laura Gómez-Consarnau<sup>1</sup>, José M. González<sup>2</sup>, Montserrat Coll-Lladó<sup>3</sup>, Pontus Gourdon<sup>4</sup>, Torbjörn Pascher<sup>5</sup>, Richard Neutze<sup>6</sup>, Carlos Pedrós-Alió<sup>3</sup> & Jarone Pinhassi<sup>1</sup>

Proteorhodopsins are bacterial light-dependent proton pumps. Their discovery within genomic material from uncultivated marine bacterioplankton caused considerable excitement because it indicated a potential phototrophic function within these organisms, which had previously been considered strictly chemotrophic<sup>1</sup>. Subsequent studies established that sequences encoding proteorhodopsin are broadly distributed throughout the world's oceans<sup>2–5</sup>. Nevertheless, the role of proteorhodopsins in native marine bacteria is still unknown<sup>6</sup>. Here we show, from an analysis of the complete genomes of three marine Flavobacteria, that cultivated bacteria in the phylum Bacteroidetes, one of the principal components of marine bacterioplankton, contain proteorhodopsin. Moreover, growth experiments in both natural and artificial seawater (low in labile organic matter, which is typical of the world's oceans) establish that exposure to light results in a marked increase in the cell yield of one such bacterium (*Dokdonia* sp. strain MED134) when compared with cells grown in darkness. Thus, our results show that the phototrophy conferred by proteorhodopsin can provide critical amounts of energy, not only for respiration and maintenance but also for active growth of marine bacterioplankton in their natural environment.

Rhodopsins are found in the domains Archaea, Bacteria and Eukarya. Rhodopsins in Archaea function as energy-transducing light-driven proton or chloride pumps, or as phototactic sensory proteins. In Eukarya, rhodopsins function primarily as sensory proteins and, for example, account for colour vision in the human retina. The recent discovery of rhodopsins in Bacteria (proteorhodopsin; PR) came after the sequence analysis of a cloned genome region from a marine bacterium of the uncultivated SAR86 clade<sup>1</sup>. Subsequent screening of DNA from different oceans revealed a very large diversity of PR in bacteria belonging to divergent clades of the Alpha-proteobacteria and Gammaproteobacteria classes<sup>3–8</sup>. Studies on reconstituted PR overproduced in *Escherichia coli* have established that it functions as a light-driven proton pump with the potential to generate energy for cell growth or maintenance<sup>1,6</sup>. However, the physiological role and benefits of PR in native Bacteria in the marine environment have not been demonstrated.

Alphaproteobacteria and Gammaproteobacteria, together with members of the Bacteroidetes phylum, are the most abundant groups of heterotrophic bacteria in the sea<sup>9–12</sup>. In the present study we examined whole-genome sequences of three bacteria belonging to the class Flavobacteria, phylum Bacteroidetes (*Dokdonia* sp. strain MED134, *Polaribacter* sp. strain MED152 and *Leeuwenhoekiella blandensis* strain MED217<sup>1</sup>). These bacteria were isolated from surface water from the Mediterranean Sea and were successfully cultured (Fig. 1).



**Figure 1 | Images of Flavobacteria isolates.** **a–c**, Colony morphology of *Dokdonia* sp. MED134 (**a**), *Polaribacter* sp. MED152 (**b**) and *L. blandensis* MED217<sup>1</sup> (**c**) growing on marine agar (Difco). **d, e**, Epifluorescence microscopy images of MED134 growing in the dark (**d**) and in the light (**e**), showing differences in cell morphology. Scale bars, 1  $\mu$ m.

<sup>1</sup>Marine Microbiology, Department of Biology and Environmental Sciences, University of Kalmar, SE-39182 Kalmar, Sweden. <sup>2</sup>Department of Microbiology and Cell Biology, University of La Laguna, ES-38206 La Laguna, Tenerife, Spain. <sup>3</sup>Institut de Ciències del Mar, CMIMA-CSIC, ES-08003 Barcelona, Catalonia, Spain. <sup>4</sup>Department of Chemical and Biological Engineering, Chalmers University of Technology, SE-41296 Göteborg, Sweden. <sup>5</sup>Chemical Physics, Kemicentrum, Lund University, SE-22100 Lund, Sweden. <sup>6</sup>Department of Chemistry, Biochemistry and Biophysics, Gothenburg University, SE-40530 Göteborg, Sweden.

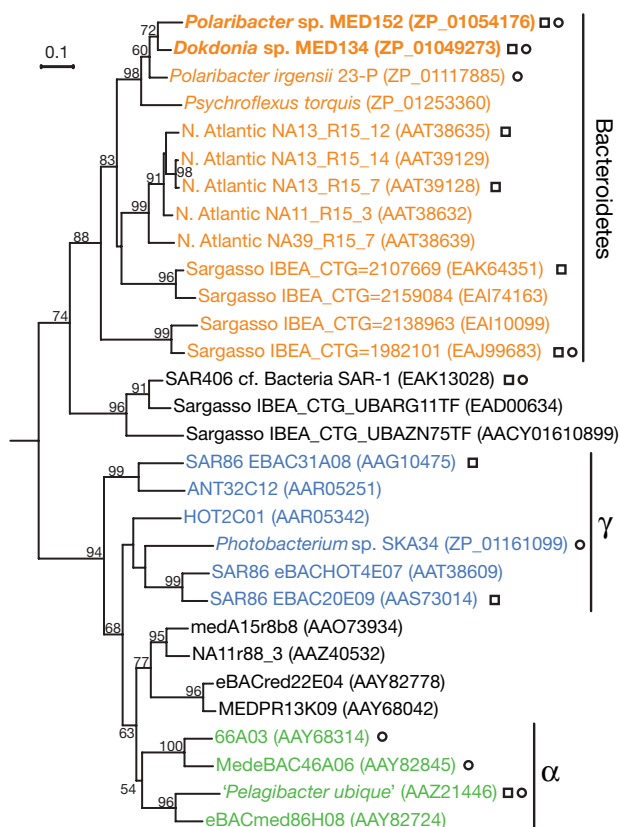
Genome analysis revealed the presence of genes encoding PR in MED134 and MED152 but not in MED217<sup>T</sup>. Furthermore, these bacteria lacked genes for bacteriochlorophyll synthesis, reaction centres or light-harvesting complexes necessary to perform photosynthesis. This indicates that, if present, phototrophy in MED134 and MED152 must be due to the activity of PR.

Proteorhodopsin genes from MED134 and MED152 encoded peptides of 247 and 243 amino acid residues, respectively, sharing a sequence similarity of 83%. They are predicted to be heptahelical integral membrane proteins, as are PRs from other marine bacteria such as the SAR86 clade and '*Pelagibacter ubique*', the first cultivated member of the abundant SAR11 clade<sup>13</sup> (Supplementary Fig. S1). Phylogenetic analysis showed that PRs from MED134 and MED152 formed a distinct cluster with two unpublished PR orthologues from Flavobacteria isolates reported in GenBank and previously unclassified environmental sequences from the North Atlantic and the Sargasso Sea<sup>5,8</sup> (Fig. 2). The similarity between PR sequences within this putative Bacteroidetes cluster ranged from 57% to 85%, whereas the similarity to sequences outside this cluster ranged from 43% to 49%. Our results therefore substantiate a previous suggestion<sup>5</sup> that some PR sequences from uncultivated bacteria belong to the phylum Bacteroidetes.

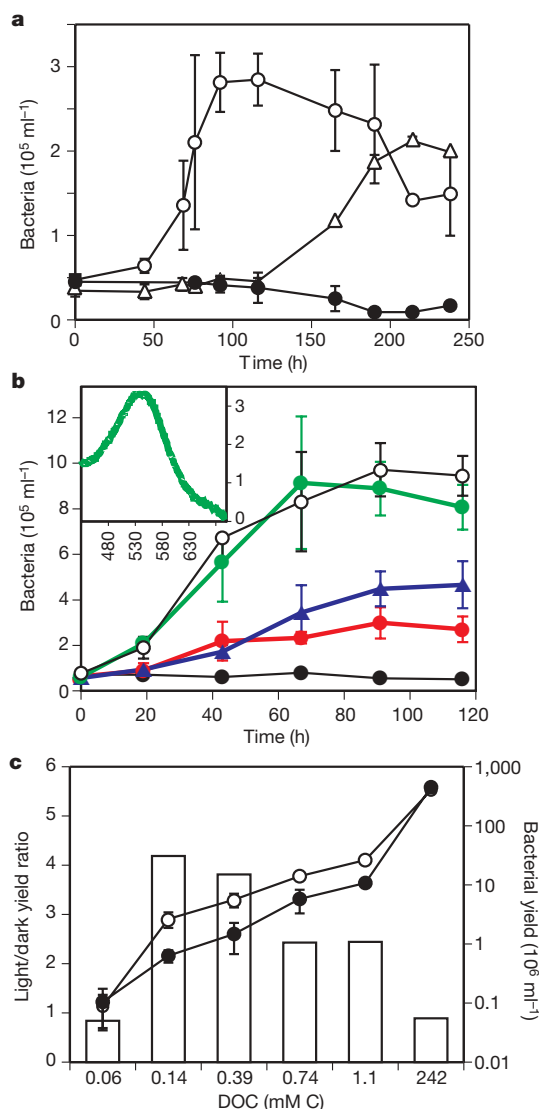
Several key amino acid positions necessary for energy generation are conserved among rhodopsins functioning as proton or ion pumps (Supplementary Fig. S1). Lys 216, which binds retinal to helix G through a protonated Schiff base in bacteriorhodopsin, was conserved as Lys 230 in MED134 and MED152. Asp 85, the proton acceptor from the Schiff base, was conserved as Asp 97. In PR from

marine bacteria Glu 108 replaces Asp 96, which facilitates Schiff-base reprotonation during the latter half of the bacteriorhodopsin photocycle<sup>14,15</sup>. Nevertheless, experiments with SAR86 and *P. ubique* PRs have demonstrated that they also function as proton pumps<sup>1,6</sup>. Thus, the essential mechanism of light-driven proton pumping by bacteriorhodopsin<sup>16</sup> seems also to apply for PR. Furthermore, laser flash (532 nm)-induced absorbance changes (probed at 500 nm) in membrane extracts from MED134 showed reaction times less than 50 ms, consistent with the presence of PR functioning as a proton pump (Supplementary Fig. S2).

A phototrophic role for PR implies that its absorption maximum would be tuned to optimize the overlap with environmental light<sup>2,7</sup>. Thus, PRs in near-surface seawater typically have leucine at amino



**Figure 2 | Phylogenetic tree of PR amino acid sequences.** The position of MED134 and MED152 (bold type) in relation to selected sequences of tentative Bacteroidetes PRs from the Sargasso Sea and the North Atlantic as well as Alphaproteobacteria ( $\alpha$ ) and Gammaproteobacteria ( $\gamma$ ). Squares and circles denote sequences aligned in Supplementary Fig. S1 and included in the gene arrangement in Supplementary Fig. S3, respectively. The numbers at nodes are bootstrap values more than 50% after 100 replicates. The scale bar represents substitutions per site.



**Figure 3 | Growth of MED134 in seawater cultures.** **a**, Cultures grown in natural sterile filtered and autoclaved seawater, exposed to light (open circles) or in darkness (filled circles). After 120 h, duplicate cultures grown in darkness were exposed to light (open triangles). **b**, Cultures exposed to white light (open circles), blue light (blue triangles), green light (green circles) and red light (red filled circles) and cultures maintained in the dark (black filled circles). Inset: absorption spectrum of MED134 PR purified from *E. coli* membranes (for details see Supplementary methods); the x axis shows wavelength (nm) and the y axis absorbance ( $10^{-2}$ ). **c**, Cell yields in artificial seawater, incubated in the light (open circles) and in the dark (filled circles), enriched with dissolved organic matter, in comparison with unenriched controls (0.06 mM C) and full-strength medium (242 mM C). Columns represent light/dark ratios of cell yield. Error bars denote s.d. for duplicate or triplicate cultures.

acid position 105 (eBAC31A08 numbering) and show absorption maxima near 530 nm (green light), whereas PRs from deeper water have glutamine at this position and absorb near 490 nm (blue light)<sup>2,7,15</sup>. Consistent with these findings is the observation that the near-surface Bacteroidetes PRs contained a hydrophobic side-chain Met 105 (similar to Leu; Supplementary Fig. S1) and showed an absorption maximum at 535 nm (Fig. 3b, inset).

Functional PR requires the covalent binding of retinal, which is synthesized from  $\beta$ -carotene. MED134 and MED152 contained the predicted genes *crtEBIY* (for gene arrangement see Supplementary Fig. S3) encoding enzymes needed to synthesize  $\beta$ -carotene from farnesyl diphosphate and isopentenyl diphosphate<sup>17</sup>. A gene encoding a candidate for Blh, an enzyme that converts  $\beta$ -carotene to retinal<sup>8</sup>, was found next to the gene encoding PR in the Bacteroidetes, whereas in Proteobacteria the *blh* gene is next to putative genes in the pathway for retinal synthesis<sup>6,8</sup> (Supplementary Fig. S3).

Since its discovery in the sea by cloning environmental DNA, the lack of cultivated bacteria that have and express PR has hampered the determination of its physiological relevance to marine bacteria *in situ*. The gene encoding PR was recently found in cultivated *P. ubique*, and mass spectroscopy indicated that *Pelagibacter* PR is abundant in the surface ocean<sup>6</sup>. Laboratory experiments with *P. ubique*, however, showed no differences in cell yield when growing in light or darkness despite the fact that PR was present in cell membranes. The authors therefore speculated that the benefits of PR might be more pronounced when bacterial growth is limited by organic matter availability. We monitored the growth of MED134 in experiments with natural seawater exposed to light or darkness. Bacteria growing in moderate light reached a maximal abundance of  $3 \times 10^5$  cells ml<sup>-1</sup> after 100 h (Fig. 3a). In contrast, bacteria remained below  $0.5 \times 10^5$  cells ml<sup>-1</sup> in the dark. When dark cultures were exposed to light after 120 h, growth was initiated and bacteria reached  $2 \times 10^5$  cells ml<sup>-1</sup> within 100 h, whereas bacteria in the cultures remaining in darkness slowly decreased in numbers. This is the first confirmation *in vivo* that light has a definite positive impact on the growth of marine bacteria possessing the gene encoding PR.

Experiments using light of different wavelengths showed that this enhanced growth was stimulated primarily by green rather than blue or red light (Fig. 3b). This correlated well with the measured absorption spectrum of MED134 PR (Fig. 3b, inset) and substantiates our hypothesis that the observed growth enhancement derives from the direct absorption of light by PR. Furthermore, analysis of the expression of the gene encoding PR with the use of reverse-transcriptase-mediated polymerase chain reaction for detecting the corresponding mRNA showed a significant upregulation of the PR mRNA in light in comparison with dark cultures (Fig. 4). These observations support the notion that PR fulfils a phototrophic function in marine bacteria.

We further investigated how different concentrations of dissolved organic matter (DOM, as peptone and yeast extract) affected the growth of MED134 (Fig. 3c). In cultures with artificial seawater without enrichment (0.06 mM C) little growth was observed in either light or darkness, whereas enrichment with DOM corres-

ponding to 0.14 or 0.39 mM C (final concentration) resulted in bacterial cell yields about fourfold higher in light cultures. Greater enrichment (0.74 or 1.10 mM C) gave yields about 2.5-fold higher in the light cultures, whereas in full-strength medium (242 mM C) similar bacterial cell yields were observed in both light and dark conditions. In addition, cells grown in light-exposed cultures seemed larger than those in dark cultures (Fig. 1d, e). Thus, exposure to light confers a stronger selective advantage when MED134 grows at low or intermediate concentrations of labile organic matter, closer to those found in marine surface waters, than in richer media. It is noteworthy that *Dokdonia* sp. MED134 thrives in conditions under which resources are abundant, whereas *P. ubique*, in which PRs have been suggested to have a subtle function<sup>6</sup>, grows only under oligotrophic conditions. Factors limiting its growth have yet to be identified<sup>6</sup>. These findings emphasize the need to clarify the intricate relationships between a PR-mediated light response, and the availability and composition of organic matter for typically oligotrophic bacteria, in comparison with bacteria adapted to eutrophic environments.

Members of the Bacteroidetes phylum in general, and Flavobacteria in particular, are important in the degradation of organic matter during and after algal blooms in the sea<sup>18–20</sup>. Our findings indicate that phototrophy may allow these organisms to maintain net growth during periods when concentrations of organic matter are declining. If this were true for other PR-bearing bacterioplankton, PR-mediated photoheterotrophy would increase their growth efficiency. An interesting consequence is that, for a given amount of DOM, bacterial assemblages dominated by species expressing genes encoding PR would produce more biomass than assemblages of bacteria lacking PR, making it available to higher trophic levels in the marine microbial food web.

Phototrophism has now been found in all main microbial groups in the sea. These include eukaryotic algae and cyanobacteria performing chlorophyll *a*-driven photosynthesis, Proteobacteria using bacteriochlorophylls for generating energy<sup>21–23</sup>, and Proteobacteria<sup>1,3</sup>—and now also Bacteroidetes—using PR. The phylogenetic diversification and wide geographic distribution of genes for PR in the world's oceans indicate that evolution has favoured organisms with the potential to complement their chemotrophic life style by phototrophy<sup>3,24–26</sup>. Surface and near-surface marine microorganisms are bathed in light and it is therefore not surprising that diverse strategies have evolved to exploit such an abundant energy source.

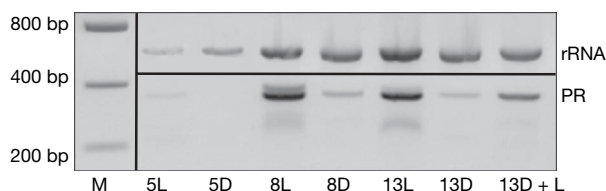
## METHODS

**Isolation of Flavobacteria strains.** Bacteria were isolated from Northwest Mediterranean Sea surface water (0.5 m depth) collected 1 km off the Catalan coast at the Blanes Bay Microbial Observatory (41° 40' N, 2° 48' E, Spain). Strains MED134 and MED152 were isolated from ZoBell agar plates inoculated on 20 March 2001, and strain MED217<sup>T</sup> on 23 May 2001.

**Sequencing, annotation and phylogenetic analysis.** Whole-genome sequencing was performed by the J. Craig Venter Institute through the Gordon and Betty Moore Foundation initiative in Marine Microbiology (<https://research.venterlinstitute.org/moore/>). PR sequences were aligned by using ClustalW, and a Neighbour-Joining phylogenetic tree was constructed (see Supplementary Methods).

**Seawater culture experiments.** For the natural seawater experiment, water from the North Sea was mixed with aged seawater from the North Atlantic. For the DOM gradient experiment, artificial seawater (35 practical salinity units, prepared from Sea Salts; Sigma) was enriched with peptone and yeast extract. All cultures were enriched with N and P to avoid inorganic nutrient limitation. Cultures were incubated under continuous light (180  $\mu$ mol photons m<sup>-2</sup> s<sup>-1</sup>) or in the dark at 20 °C. Dissolved organic carbon was measured with a high-temperature carbon analyser. Bacteria were counted by epifluorescence microscopy. For details see Supplementary Methods.

**Expression analysis.** RNA were extracted and purified from cell pellets with RNeasy Lysis Buffer (Qiagen). mRNA was reverse transcribed with the two-step protocol of RETROscript Kit (Ambion Inc.) and a MED134 PR-specific primer set. For details see Supplementary Methods.



**Figure 4 | RT-PCR analysis of PR expression in MED134 compared with 16S rRNA levels.** PR was expressed preferentially in the light. Bacteria were grown in natural sterile filtered and autoclaved seawater exposed to light or in the dark. Numbers indicate days since inoculation; D, dark incubation; L, light incubation. 13D + L denotes a sample incubated for 10 days in the dark followed by 3 days in the light. M, DNA size marker; bp, base pairs.



Received 24 September; accepted 25 October 2006.

1. Bèjà, O. *et al.* Bacterial rhodopsin: evidence for a new type of phototrophy in the sea. *Science* **289**, 1902–1906 (2000).
2. Bèjà, O., Spudich, E. N., Spudich, J. L., Leclerc, M. & DeLong, E. F. Proteorhodopsin phototrophy in the ocean. *Nature* **411**, 786–789 (2001).
3. de la Torre, J. R. *et al.* Proteorhodopsin genes are distributed among divergent marine bacterial taxa. *Proc. Natl Acad. Sci. USA* **100**, 12830–12835 (2003).
4. Sabehi, G. *et al.* Novel proteorhodopsin variants from the Mediterranean and Red Seas. *Environ. Microbiol.* **5**, 842–849 (2003).
5. Venter, J. C. *et al.* Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304**, 66–74 (2004).
6. Giovannoni, S. J. *et al.* Proteorhodopsin in the ubiquitous marine bacterium SAR11. *Nature* **438**, 82–85 (2005).
7. Man, D. *et al.* Diversification and spectral tuning in marine proteorhodopsins. *EMBO J.* **22**, 1725–1731 (2003).
8. Sabehi, G. *et al.* New insights into metabolic properties of marine bacteria encoding proteorhodopsins. *PLoS Biol.* **3**, e273 (2005).
9. Giovannoni, S. & Rappé, M. in *Microbial Ecology of the Oceans* (ed. Kirchman, D.) 47–84 (Wiley-Liss, New York, 2000).
10. Glöckner, F. O., Fuchs, B. M. & Amann, R. Bacterioplankton composition of lakes and oceans: a first comparison based on fluorescence in situ hybridization. *Appl. Environ. Microbiol.* **65**, 3721–3726 (1999).
11. Hagström, Å. *et al.* Use of 16S ribosomal DNA for delineation of marine bacterioplankton species. *Appl. Environ. Microbiol.* **68**, 3628–3633 (2002).
12. Kirchman, D. L. The ecology of Cytophaga–Flavobacteria in aquatic environments. *FEMS Microbiol. Ecol.* **39**, 91–100 (2002).
13. Rappé, M. S., Connon, S. A., Vergin, K. L. & Giovannoni, S. J. Cultivation of the ubiquitous SAR11 marine bacterioplankton clade. *Nature* **418**, 630–633 (2002).
14. Dioumaev, A. K. *et al.* Proton transfers in the photochemical reaction cycle of proteorhodopsin. *Biochemistry* **41**, 5348–5358 (2002).
15. Wang, W.-W., Sineshchekov, O. A., Spudich, E. N. & Spudich, J. L. Spectroscopic and photochemical characterization of a deep ocean proteorhodopsin. *J. Biol. Chem.* **278**, 33985–33991 (2003).
16. Neutze, R. *et al.* Bacteriorhodopsin: a high-resolution structural view of vectorial proton transport. *Biochim. Biophys. Acta* **1565**, 144–167 (2002).
17. Teramoto, M., Takaichi, S., Inomata, Y., Ikenaga, H. & Misawa, N. Structural and functional analysis of a lycopene L-monocyclase gene isolated from a unique marine bacterium that produces myxol. *FEBS Lett.* **545**, 120–126 (2003).
18. Abell, G. C. J. & Bowman, J. P. Ecological and biogeographic relationships of class Flavobacteria in the Southern Ocean. *FEMS Microbiol. Ecol.* **51**, 265–277 (2005).
19. Pinhassi, J. *et al.* Changes in bacterioplankton composition under different phytoplankton regimens. *Appl. Environ. Microbiol.* **70**, 6753–6766 (2004).
20. Riemann, L., Steward, G. F. & Azam, F. Dynamics of bacterial community composition and activity during a mesocosm diatom bloom. *Appl. Environ. Microbiol.* **66**, 578–587 (2000).
21. Shiba, T., Simidu, U. & Taga, N. Distribution of aerobic bacteria which contain bacteriochlorophyll *a*. *Appl. Environ. Microbiol.* **38**, 43–48 (1979).
22. Kolber, Z. S., Van Dover, C. L., Niederman, R. A. & Falkowski, P. G. Bacterial photosynthesis in surface waters of the open ocean. *Nature* **407**, 177–179 (2000).
23. Bèjà, O. *et al.* Unsuspected diversity among marine aerobic anoxygenic phototrophs. *Nature* **415**, 630–633 (2002).
24. Bielawski, J. P., Dunn, K. A., Sabehi, G. & Bèjà, O. Darwinian adaptation of proteorhodopsin to different light intensities in the marine environment. *Proc. Natl Acad. Sci. USA* **101**, 14824–14829 (2004).
25. Frigaard, N.-U., Martinez, A., Mincer, T. J. & DeLong, E. F. Proteorhodopsin lateral gene transfer between marine planktonic Bacteria and Archaea. *Nature* **439**, 847–850 (2006).
26. Sabehi, G., Bèjà, O., Suzuki, M. T., Preston, C. M. & DeLong, E. F. Different SAR86 subgroups harbour divergent proteorhodopsins. *Environ. Microbiol.* **6**, 903–910 (2004).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank S. Arnautovic, J. O. Ekström, M. Widell, E. Lundberg and E. Lindehoff for help with growth experiments, ultracentrifugation, cloning, dissolved organic carbon and nutrient analysis, respectively, and T. Berman for helpful comments on the manuscript. We thank the Swedish Science Council, the Spanish Ministerio de Educación y Ciencia, Swegene, EMEP, and SSF for supporting this research.

**Author Information** The genomes of strains MED134, MED152 and MED217 are deposited in GenBank under accession numbers AAMZ000000000, AANA000000000 and AANC000000000, and their 16S rRNA gene sequences under accession numbers DQ481462, DQ481463 and DQ294290, respectively. The amino acid sequences of MED134 and MED152 PR are deposited in GenBank under accession numbers ZP\_01049273 and ZP\_01054176. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.P. ([jarone.pinhassi@hik.se](mailto:jarone.pinhassi@hik.se)).

## LETTERS

# Histone arginine methylation regulates pluripotency in the early mouse embryo

Maria-Elena Torres-Padilla<sup>1</sup>, David-Emlyn Parfitt<sup>1</sup>, Tony Kouzarides<sup>1</sup> & Magdalena Zernicka-Goetz<sup>1</sup>

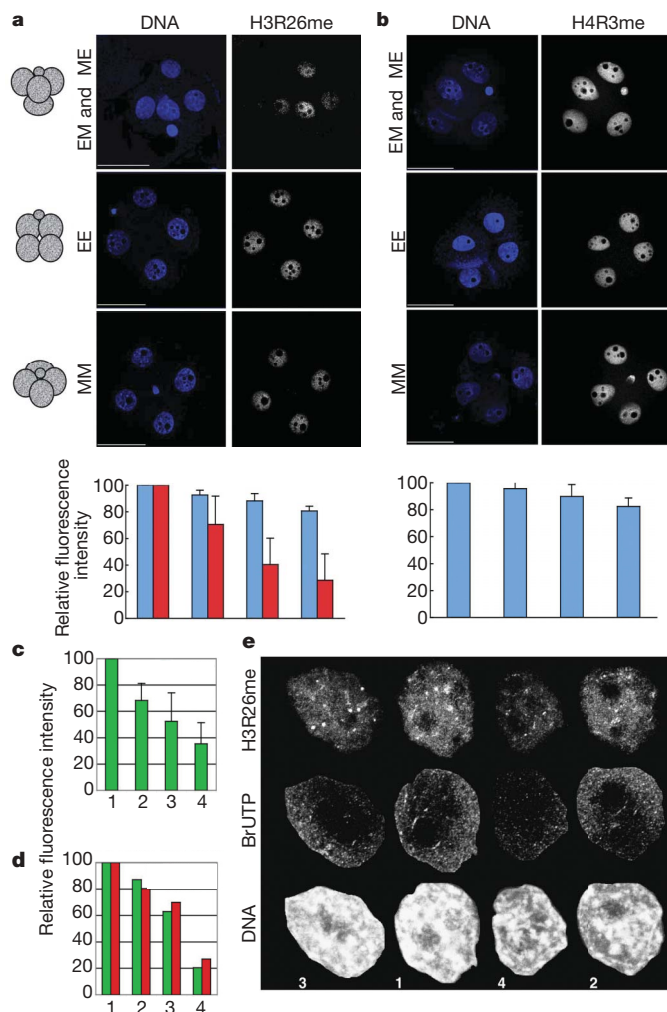
It has been generally accepted that the mammalian embryo starts its development with all cells identical, and only when inside and outside cells form do differences between cells first emerge. However, recent findings show that cells in the mouse embryo can differ in their developmental fate and potency as early as the four-cell stage<sup>1–4</sup>. These differences depend on the orientation and order of the cleavage divisions that generated them<sup>2,5</sup>. Because epigenetic marks are suggested to be involved in sustaining pluripotency<sup>6,7</sup>, we considered that such developmental properties might be achieved through epigenetic mechanisms. Here we show that modification of histone H3, through the methylation of specific arginine residues, is correlated with cell fate and potency. Levels of H3 methylation at specific arginine residues are maximal in four-cell blastomeres that will contribute to the inner cell mass (ICM) and polar trophoctoderm and undertake full development when combined together in chimaeras. Arginine methylation of H3 is minimal in cells whose progeny contributes more to the mural trophoctoderm and that show compromised development when combined in chimaeras. This suggests that higher levels of H3 arginine methylation predispose blastomeres to contribute to the pluripotent cells of the ICM. We confirm this prediction by overexpressing the H3-specific arginine methyltransferase CARM1 in individual blastomeres and show that this directs their progeny to the ICM and results in a dramatic upregulation of *Nanog* and *Sox2*. Thus, our results identify specific histone modifications as the earliest known epigenetic marker contributing to development of ICM and show that manipulation of epigenetic information influences cell fate determination.

To address whether epigenetic differences exist between blastomeres at the four-cell stage, we focused on histone methylation marks related to transcriptional activation<sup>8–10</sup>. Because the divisions of two-cell-stage blastomeres differ in orientation in relation to the animal–vegetal axis of the egg<sup>5,11</sup>, the shapes of four-cell embryos vary: blastomeres fill the apices of a tetrahedron when they undergo one equatorial (E) and one meridional (M) division, or they lie on a similar plane when they undergo either two E or two M divisions (Fig. 1a). Although any combination of the temporal sequence of such divisions is possible, sequential M and E divisions are most common (about 80%), but they can occur in either order. We found that whereas embryos that underwent either two E or two M divisions did not show significant variations in the levels of H3 methylation at Arg 26 (H3R26me) between four-cell-stage blastomeres (Fig. 1a, EE and MM embryos; blue bars), tetrahedral embryos showed marked differences in H3R26me levels between their blastomeres (Fig. 1a, EM and ME embryos; red bars). In the latter, the weakest level of H3R26me was generally less than 40% of the cell giving the strongest signal ( $P = 0.0002$ ). Measuring the intensity of DNA staining indicated that the variation of H3R26me levels was not related to differences in the content of DNA resulting from replication (not

shown). We confirmed that this variation in H3R26me levels did not result from the confocal scanning of embryos of differing shapes by scanning individual cells of disaggregated embryos, where we found a similar outcome (Supplementary Fig. S1). In addressing whether the methylation of other arginine residues also showed differences between four-cell blastomeres, we found that levels of H3R2me and H3R17me (Supplementary Figs S2 and S3a) varied between blastomeres correlating with embryo morphology in a similar way to H3R26me. This is consistent with Arg 2 and Arg 17 being targets for the same methyltransferase, CARM1, as Arg 26 (refs 12, 13). When we analysed CARM1 distribution in four-cell blastomeres, we found that CARM1 levels varied with the same tendency as those of H3R26me (Supplementary Fig. S4b, c). In contrast, methylation of H4R3, which is the target of a different methyltransferase, protein arginine methyltransferase 1 (PRMT1; refs 14, 15), seemed equivalent between blastomeres regardless of embryo morphology (Fig. 1b). Thus, the differences in the levels of histone arginine methylation in four-cell-stage blastomeres are specific. We also observed that levels of BrUTP incorporation in late four-cell-stage embryos were highest in the blastomeres that were enriched for H3R26me, indicative of elevated levels of global transcription in these cells (Fig. 1c–e). Because H3R26me showed the biggest difference in its distribution between blastomeres (Supplementary Fig. S3b), we concentrated on analysing this modification further.

Because different developmental fate and potential can be ascribed to blastomeres depending on the orientation and the order of division from the two-cell to the four-cell stage in relation to the animal–vegetal axis, we wished to determine whether differences in H3R26me related to patterns of division. To this end we first grouped embryos according to their cleavage patterns to the four-cell stage. When the earlier of the second cleavages is M and the later E (ME embryos), the earlier dividing two-cell blastomere contributes most of its progeny to the embryonic (ICM and adjacent polar trophoctoderm) and the later one to the abembryonic part (ICM and mural trophoctoderm) of the blastocyst. When the earlier second cleavage is E and the later is M (EM embryos), the earlier blastomere can give rise to either the embryonic or abembryonic part of the blastocyst. By contrast, when second cleavage divisions are of similar orientation (MM or EE), the allocation of blastomere progeny is random. The ME group of embryos thus allowed the identification of individual four-cell blastomeres that have a predictable fate within the blastocyst<sup>5</sup>. Moreover, blastomeres resulting from the E division that inherit the ‘vegetal’ cytoplasm in ME embryos tend to contribute more to the mural trophoctoderm and do not complete development when combined with the same type of ‘vegetal’ cells in chimaeric embryos<sup>2</sup>. In contrast, chimaeras of blastomeres that arise from early M divisions (with ‘animal–vegetal’ cytoplasm) could complete development with full success, and chimaeras constructed only from ‘animal’ blastomeres could also develop although with reduced success.

<sup>1</sup>The Wellcome Trust/Cancer Research UK Gurdon Institute, University of Cambridge, Tennis Court Road, Cambridge CB2 1QN, UK.

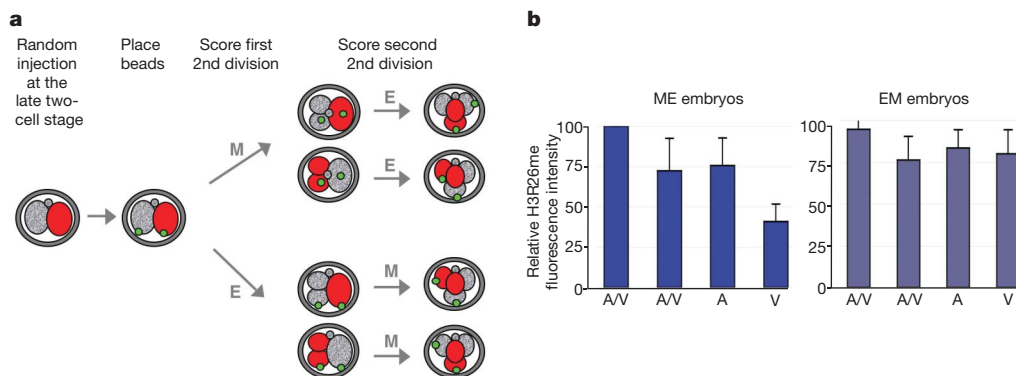


To distinguish the progeny of two-cell-stage blastomeres, we injected one two-cell blastomere at random with rhodamine-dextran and then monitored the second cleavage divisions (Supplementary Fig. S5a). Immunostaining then revealed similar levels of H3R26me in the progeny of both E and M divisions of EM embryos (compare M cells with E cells;  $P = 0.112$ ; Supplementary Fig. S5b, c). In ME

**Figure 1 | Levels of H3R26me are different in blastomeres of four-cell-stage embryos and correlate with their spatial arrangement.** **a**, Four-cell-stage embryos were stained for H3R26me and grouped according to their shape: tetrahedral (EM and ME), EE (flattened, polar body on one side) or MM (flattened, polar body in the middle). Shown are projections, including all sections, of representative embryos. Fluorescence levels were quantified and normalized against the blastomere showing the highest level, which was set at 100%. Decreasing values of fluorescence were normalized and averaged accordingly ( $n = 18$ ). Each bar represents the relative fluorescence level of each of the four blastomeres: red bars, EM and ME embryos; blue bars, MM or EE embryos. Scale bar, 50  $\mu\text{m}$ . **b**, Differences in histone arginine methylation levels in four-cell-stage blastomeres are specific: only residues that are CARM1 targets, and not PRMT1, show differential distribution. The histogram gives results for embryos of all shapes. **c**, Four-cell-stage blastomeres show different global transcriptional activities. BrUTP incorporation was measured in sections from nuclei of four-cell-stage embryos captured every 0.6  $\mu\text{m}$  ( $n = 12$ ). Projections were used after cropping off the nuclei with the Velocity software to quantify active (nuclear) transcription. Values were normalized as in **a**. **d**, **e**, Global transcription levels correlate with global H3R26me levels. **d**, Quantification of BrUTP incorporation (green) and H3R26me (red) of nuclei of each four-cell blastomere of a representative embryo. **e**, Nuclei shown at the same scale, numbers at the bottom correspond to the blastomere numbers of the graph in **d**. Error bars indicate s.d.

embryos, however, the M sister cells had significantly higher levels of H3R26 methylation than both E sisters, with one of the E-derived blastomeres showing significantly less methylation than either of the M-derived blastomeres ( $P < 0.0001$ ; Supplementary Fig. S5b, c). In confirmation of our earlier observations (Fig. 1), EE and MM embryos did not show differences in their relative H3R26me levels. Thus, in ME embryos in which prediction of developmental properties is possible, higher levels of H3R26me are seen in blastomeres expected to contribute more cells to the embryonic part of the blastocyst and lower levels in blastomeres expected to contribute to the abembryonic part.

Given the reported differences in developmental potential of ‘animal’ and ‘vegetal’ blastomeres arising from E divisions in ME embryos<sup>2</sup>, we next examined whether these cells differed in their H3R26me levels. To this end we randomly injected a two-cell-stage blastomere with rhodamine-dextran as before and then applied green fluorescent beads to the vegetal membranes of the two blastomeres as a second label (Fig. 2a and ref. 2). The rhodamine label allowed us to score the order and the plane of division to the four-cell-stage and the beads served as ‘vegetal’ markers. We found that in ME embryos, the ‘vegetal’ blastomere always had significantly lower levels of H3R26me than the ‘animal’ or ‘animal/vegetal’ blastomeres ( $P < 0.0001$ ;



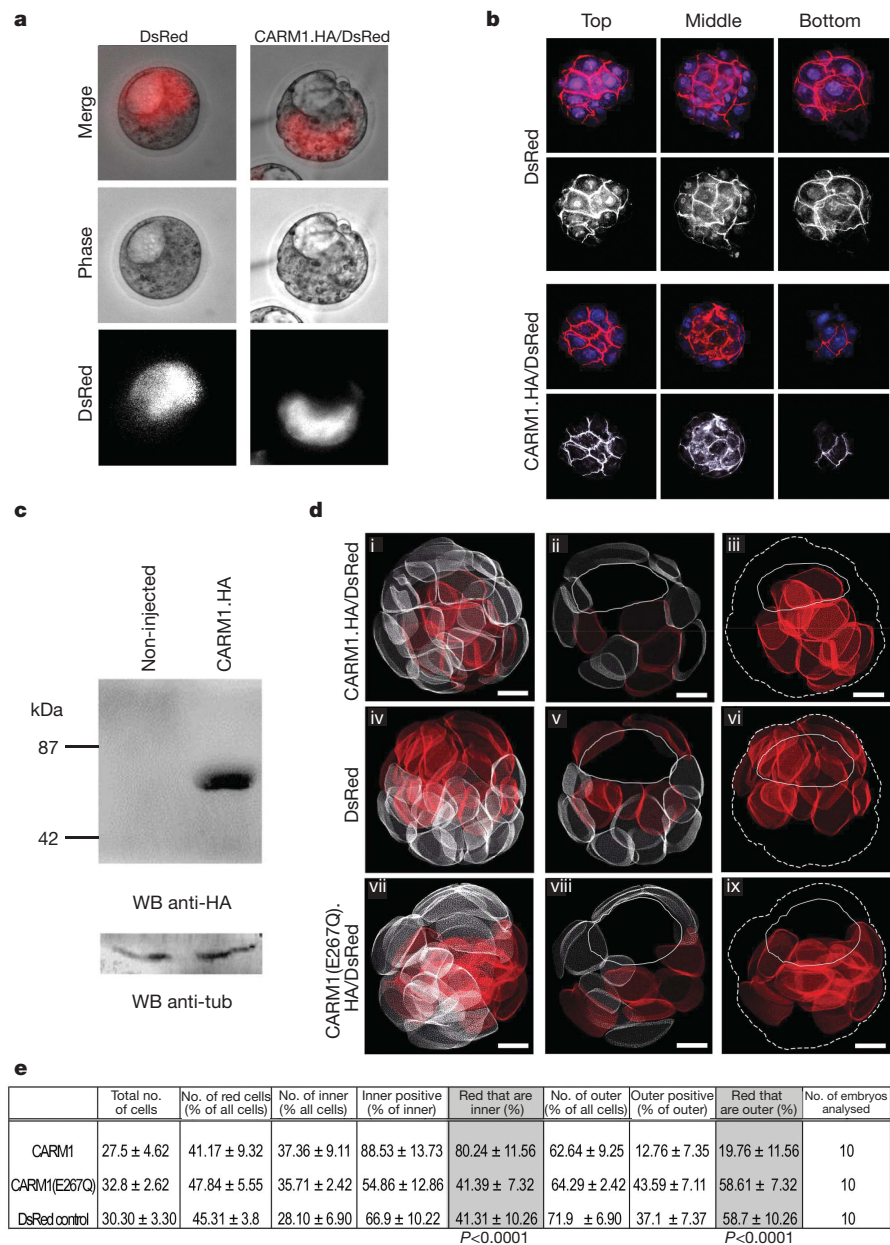
**Figure 2 | The ‘vegetal’ blastomere in ME embryos has the lowest levels of H3R26me.** **a**, Design to determine the identity of four-cell-stage blastomeres according to division orientation, order, and blastomere positioning. A two-cell-stage blastomere was microinjected with rhodamine-dextran. We then placed a green fluorescent bead in the ‘vegetal’ membrane of the two blastomeres. Divisions were scored and embryos were stained for H3R26me at the late four-cell stage. The position of the bead and the presence of rhodamine allowed the identification of the blastomeres as animal/vegetal

(A/V, derived from M divisions), or animal or vegetal (A or V, derived from E divisions) in EM ( $n = 10$ ) or ME ( $n = 9$ ) embryos. **b**, The vegetal blastomere of ME embryos shows the lowest levels of H3R26me, whereas the vegetal blastomere of EM embryos shows similar levels to the animal or animal/vegetal blastomeres. H3R26me levels were quantified as in Fig. 1. Because there are two A/V blastomeres after one M division there are two columns labelled A/V. Error bars indicate s.d.



Fig. 2b). In contrast, in EM embryos the ‘animal’ and ‘vegetal’ blastomeres from the E division had equivalent levels of H3R26me (Fig. 2b). Taken together, our results indicate that more extensive H3 arginine methylation in M blastomeres of ME embryos is correlated with their greater contribution to the embryonic part of the blastocyst. In contrast, blastomeres having the lowest levels of H3

arginine methylation (that is, ‘vegetal’ blastomeres) are those predicted to contribute mainly to the abembryonic part. Because levels of methylation at Arg 2, Arg 17 and Arg 26 varied similarly and because these three residues are the specific targets of CARM1/PRMT4, which is expressed maternally in mouse embryos (Supplementary Figs S4 and S6), we wondered whether CARM1



**Figure 3 | CARM1 overexpression in a two-cell blastomere results in the contribution of that cell predominantly to the ICM.** A late two-cell-stage blastomere was injected with mRNA for DsRed alone (control) or in combination with mRNA for CARM1.HA (HA stands for haemagglutinin). This resulted in CARM1 overexpression from the mid-four-cell stage because CARM1.HA/DsRed expression starts 6–8 h after injection (not shown). Embryos were cultured until the blastocyst stage and observed under fluorescence microscopy. DsRed was used as a lineage tracer. **a**, Representative embryos derived from DsRed only ( $n = 17$ ) or DsRed/CARM1.HA-overexpression experiments ( $n = 35$ ). **b**, Blastocysts were stained with phalloidin–Texas red and TOTO-3 (to reveal cell membranes and DNA, respectively) and analysed under confocal microscopy. Representative top, middle and bottom sections are shown. DNA is in blue; phalloidin (red) can be distinguished from DsRed because the latter is exclusively cytoplasmic. The red channel is shown as a greyscale. The progeny of the CARM1-overexpressing blastomere is predominantly within the inner cells of the

blastocyst. **c**, Overexpression of CARM1 was verified by western blot (WB) analysis in zygotes injected with mRNA for CARM1.HA/DsRed. **d**, Representative three-dimensional reconstructions of blastocysts in which mRNA for CARM1.HA/DsRed (i), DsRed only (iv) or CARM1(E267Q).HA/DsRed (vii) was microinjected at the two-cell stage. Blastocysts were stained as in **b**. Confocal z-stacks were taken at 1- $\mu$ m intervals. IMARIS software was used to outline cell membranes to create three-dimensional models of all cells of the embryo. Cells were then scored according to their position: cells completely surrounded by others are denoted as inner, those with an outer surface as outer. Cells were scored as either positive or negative for DsRed. Progeny of injected blastomere is shown in red. A middle slice is shown in ii, v and viii, where the cavity is depicted with a line. In iii, vi and ix, only the progeny of the injected blastomere is shown; the contour of the embryo is indicated by a dashed line and the position of the cavity by a solid line. Scale bar, 10  $\mu$ m. **e**, Analysis of the distribution of the progeny of CARM1- and DsRed-overexpressing blastomere at the blastocyst stage. Results are means  $\pm$  s.d.

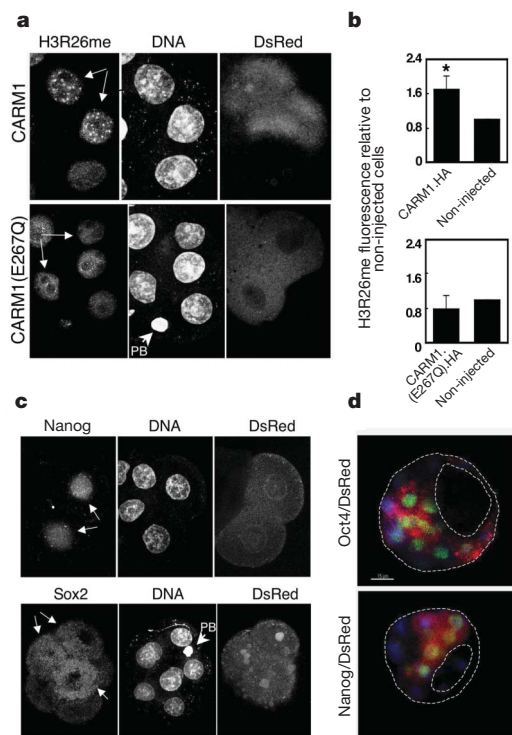
might have a function in directing developmental fate and potency. To test this, we injected CARM1 mRNA into single late two-cell-stage blastomeres aiming to elevate levels of arginine methylation in H3 in the progeny of these cells from the mid-four-cell stage. We followed the effect on cell fate by co-injecting mRNA for DsRed as a lineage tracer (Fig. 3a, b). Strikingly, the labelled clone was located in the embryonic part of the blastocyst in 31 out of 35 embryos (89%), and in no embryos were labelled cells found exclusively in the abembryonic part. We randomly selected ten of these embryos for reconstruction in three dimensions to locate every cell and determine to which lineage (ICM or trophectoderm) labelled cells had contributed. This showed that the ICM comprised 37% of all cells at this stage and that, on average, 88.5% of ICM cells were derived from the blastomere in

which CARM1 had been overexpressed (Fig. 3d, e, and Supplementary Table S1). Interestingly, in half of the embryos, all of the ICM comprised the exclusive progeny of the CARM1-overexpressing blastomere. By contrast, even though most (63%) of the blastocyst cells are outer cells, the proportion of outer cells derived from the CARM1-overexpressing blastomere was only 12% (Fig. 3e, and Supplementary Table S1). This was in contrast with control embryos, injected with mRNA for DsRed only, in which we found that 41% and 59% of cells labelled with the lineage tracer were inner and outer cells, respectively ( $P < 0.0001$ ; Fig. 3e, and Supplementary Table S2). We therefore conclude that forced overexpression of CARM1 in a two-cell blastomere leads that cell to contribute predominantly (if not exclusively) to the ICM. To establish whether the methyltransferase activity of CARM1 was required for this effect, we injected mRNA for CARM1 containing a point mutation (E267Q) and devoid of catalytic activity<sup>16</sup> into single two-cell blastomeres as above. We found that the progeny of the labelled cells could contribute to both embryonic and abembryonic regions of the blastocyst. The distribution of the progeny of the blastomere injected with CARM1(E267Q) was similar to that of the DsRed control (Fig. 3d, e, and Supplementary Table S3). The progeny of the CARM1-overexpressing blastomere showed increased levels of H3R26me, whereas blastomeres overexpressing CARM1(E267Q) did not (Fig. 4a, b). Hence, the ability of CARM1 to direct the progeny of a blastomere towards the ICM is strictly dependent on its methyltransferase activity.

We next assessed expression levels of transcription factors known to influence the development of ICM cells. We found that overexpression of CARM1 led to an early and marked upregulation of Nanog in the injected blastomeres, suggesting that the Nanog promoter is regulated by arginine methylation (Fig. 4c). In contrast, *Cdx2*, a trophectoderm marker, showed no induction after overexpression of CARM1 (not shown). We also detected increased levels of Sox2 in the progeny of the CARM1-injected blastomere (Fig. 4c). However, Oct4/Pou5f1 showed variable levels of expression in non-injected blastomeres as well as in cells overexpressing CARM1 (Supplementary Fig. S7). Note that, in contrast with Nanog, both Oct4/Pou5f1 and Sox2 were also present in the non-injected blastomeres, reflecting an earlier expression and/or their maternal inheritance<sup>17,18</sup>. The co-expression by the blastocyst stage of the ICM markers Oct4/Pou5f1 and Nanog in the progeny of the CARM1-overexpressing blastomere is consistent with the observed change in cell fate (Fig. 4d). Hence, by manipulating epigenetic information through the overexpression of a histone modifier it is possible to direct cells towards the ICM.

Our findings, in control experiments, that either of the two-cell blastomeres normally contributes its progeny to both inner and outer cells of the blastocyst are in accordance with our earlier findings<sup>1,2,4,5</sup>. However, they contrast with a report that shows that as a consequence of differential *Cdx2* levels between two-cell blastomeres, one blastomere contributes exclusively to outer cells, which normally express *Cdx2* (ref. 19). Unlike those authors we do not observe expression of *Cdx2* at the two-cell stage.

Our study demonstrates that epigenetic differences develop between blastomeres by the four-cell stage. We cannot exclude the possibility that CARM1 mediates some of its effects through targets other than histone H3. However, the differential cellular levels of H3 methylation within four-cell blastomeres can account for their different cell fates and potencies<sup>1,4,5</sup>. Cells with more extensive H3 arginine methylation are destined to contribute pluripotent progeny to the blastocyst. These cells show increased levels of transcription, including the expression of a select set of genes responsible for maintaining pluripotency such as Nanog and Sox2 (refs 18, 20, 21). An enrichment in modifications characteristic of euchromatin may have additional effects on the chromatin to generate an 'open' configuration that could sustain pluripotency in the embryo, as has been suggested for ES cells<sup>22</sup>. Indeed, the relative importance of potential changes in chromatin structure in relation to cell plasticity demands further study.



**Figure 4 | Overexpression of CARM1 results in elevated levels of arginine methylation and upregulation of Nanog and Sox2.** **a**, A two-cell-stage blastomere was injected with mRNA for CARM1.HA/DsRed. Embryos were cultured to the eight-cell stage and stained for H3R26me. Shown are three nuclei of cells from a representative embryo. The progeny of the injected blastomere is indicated by arrows (note the presence of DsRed). For overexpression of the CARM1(E267Q).HA mutant, five cells from the same embryo are shown ( $n = 6$ ). PB, polar body. **b**, Quantification of H3R26me levels in cells overexpressing CARM1 were normalized against those of non-injected cells within the same embryo (asterisk,  $P = 0.0006$ ;  $n = 14$ ). Bottom: data derived from overexpression of CARM1(E267Q).HA. Error bars indicate s.d. **c**, Embryos were injected as in **a** and stained with an anti-Nanog ( $n = 5$ ) or an anti-Sox2 ( $n = 11$ ) antibody between the six-cell and the eight-cell stage. For Nanog, four nuclei from the same embryo, two of them deriving from the CARM1-overexpressing blastomere, are shown (white arrows; note the presence of DsRed). Nanog is detectable only in the blastomeres deriving from the two-cell-stage blastomere injected with CARM1 mRNA. For Sox2, a representative embryo is shown. Note that Sox2 is mainly cytoplasmic at this stage<sup>18</sup>. We were unable to address CARM1 function by the converse experiment by RNA-mediated interference because the protein is provided maternally and its mRNA is rapidly downregulated after fertilization. However, treatment of zygotes with specific arginine methyltransferase inhibitors<sup>25</sup> showed that decreasing the level of histone arginine methylation impaired development (Supplementary Fig. S8). **d**, The progeny of a CARM1-overexpressing blastomere expresses ICM markers in the blastocyst. Blastocysts were stained for Oct4/Pou5f1 ( $n = 7$ ) or Nanog ( $n = 3$ ; green). The presence of DsRed indicates the progeny of the injected blastomere. DNA is shown in blue.

## METHODS

Embryos were collected from F<sub>1</sub> crosses (C57BL/6XCBA/H). To monitor the division from the two-cell stage, one blastomere was microinjected randomly with dextran-tetramethylrhodamine (3 kDa; Molecular Probes). Green fluorescent beads were placed in the membrane of blastomeres with the use of a Piezo driller. Embryos were observed every 20 min to determine the plane and order of division<sup>5</sup>. Immunostaining and BrUTP labelling were performed as described<sup>23</sup>. Anti-H3 asymmetric-dimethyl-Arg2, anti-H3 dimethyl-Arg17, anti-H3 dimethyl-Arg26 and anti-H4 symmetric-dimethyl-Arg3 antibodies were from Abcam (Supplementary Fig. S9); anti-Oct4 was from R&D Systems; and anti-Nanog and anti-Sox2 were from Santa Cruz. For analysis of four-cell embryos, confocal sections were taken every 0.8 µm through the whole embryo and the fluorescence signal was measured in projections with the Volocity software (Improvision). For three-dimensional reconstructions, blastocysts were stained with phalloidin–Texas red and TOTO-3. Confocal sections were captured every 1 µm and processed with IMARIS (Bitplane) and 3DVirtual Embryo software<sup>24</sup>.

Received 9 August; accepted 17 November 2006.

- Piotrowska, K., Wianny, F., Pedersen, R. A. & Zernicka-Goetz, M. Blastomeres arising from the first cleavage division have distinguishable fates in normal mouse development. *Development* **128**, 3739–3748 (2001).
- Piotrowska-Nitsche, K., Perea-Gomez, A., Haraguchi, S. & Zernicka-Goetz, M. Four-cell stage mouse blastomeres have different developmental properties. *Development* **132**, 479–490 (2005).
- Gardner, R. L. Specification of embryonic axes begins before cleavage in normal mouse development. *Development* **128**, 839–847 (2001).
- Fujimori, T., Kurotaki, Y., Miyazaki, J. & Nabeshima, Y. Analysis of cell lineage in two- and four-cell mouse embryos. *Development* **130**, 5113–5122 (2003).
- Piotrowska-Nitsche, K. & Zernicka-Goetz, M. Spatial arrangement of individual 4-cell stage blastomeres and the order in which they are generated correlate with blastocyst pattern in the mouse embryo. *Mech. Dev.* **122**, 487–500 (2005).
- Li, E. Chromatin modification and epigenetic reprogramming in mammalian development. *Nature Rev. Genet.* **3**, 662–673 (2002).
- Morgan, H. D., Santos, F., Green, K., Dean, W. & Reik, W. Epigenetic reprogramming in mammals. *Hum. Mol. Genet.* **14** (Spec. Iss. 1), R47–R58 (2005).
- Chen, D. *et al.* Regulation of transcription by a protein methyltransferase. *Science* **284**, 2174–2177 (1999).
- Ma, H. *et al.* Hormone-dependent, CARM1-directed, arginine-specific methylation of histone H3 on a steroid-regulated promoter. *Curr. Biol.* **11**, 1981–1985 (2001).
- Bauer, U. M., Daujat, S., Nielsen, S. J., Nightingale, K. & Kouzarides, T. Methylation at arginine 17 of histone H3 is linked to gene activation. *EMBO Rep.* **3**, 39–44 (2002).
- Gardner, R. L. Experimental analysis of second cleavage in the mouse. *Hum. Reprod.* **17**, 3178–3189 (2002).
- Bannister, A. J. & Kouzarides, T. Reversing histone methylation. *Nature* **436**, 1103–1106 (2005).
- Schurter, B. T. *et al.* Methylation of histone H3 by coactivator-associated arginine methyltransferase 1. *Biochemistry* **40**, 5747–5756 (2001).
- Wang, H. *et al.* Methylation of histone H4 at arginine 3 facilitating transcriptional activation by nuclear hormone receptor. *Science* **293**, 853–857 (2001).
- Strahl, B. D. *et al.* Methylation of histone H4 at arginine 3 occurs *in vivo* and is mediated by the nuclear receptor coactivator PRMT1. *Curr. Biol.* **11**, 996–1000 (2001).
- Lee, Y. H., Koh, S. S., Zhang, X., Cheng, X. & Stallcup, M. R. Synergy among nuclear receptor coactivators: selective requirement for protein methyltransferase and acetyltransferase activities. *Mol. Cell. Biol.* **22**, 3621–3632 (2002).
- Scholer, H. R., Hatzopoulos, A. K., Balling, R., Suzuki, N. & Gruss, P. A family of octamer-specific proteins present during mouse embryogenesis: evidence for germline-specific expression of an Oct factor. *EMBO J.* **8**, 2543–2550 (1989).
- Avilion, A. A. *et al.* Multipotent cell lineages in early mouse development depend on SOX2 function. *Genes Dev.* **17**, 126–140 (2003).
- Deb, K., Sivaguru, M., Yong, H. Y. & Roberts, R. M. Cdx2 gene expression and trophectoderm lineage specification in mouse embryos. *Science* **311**, 992–996 (2006).
- Mitsui, K. *et al.* The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. *Cell* **113**, 631–642 (2003).
- Chambers, I. *et al.* Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. *Cell* **113**, 643–655 (2003).
- Meshorer, E. *et al.* Hyperdynamic plasticity of chromatin proteins in pluripotent embryonic stem cells. *Dev. Cell* **10**, 105–116 (2006).
- Torres-Padilla, M. E. & Zernicka-Goetz, M. Role of TIF1α as a modulator of embryonic transcription in the mouse zygote. *J. Cell Biol.* **174**, 329–338 (2006).
- Tassy, O., Daian, F., Hudson, C., Bertrand, V. & Lemaire, P. A quantitative approach to the study of cell shapes and interactions during early chordate embryogenesis. *Curr. Biol.* **16**, 345–358 (2006).
- Cheng, D. *et al.* Small molecule regulators of protein arginine methyltransferases. *J. Biol. Chem.* **279**, 23892–23899 (2004).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank P. Greda for bead labelling, C. Lee for assistance, D. Glover for comments on the manuscript, M. Stallcup for the CARM1 expression vectors, and M. Bedford for providing the *Carm1*<sup>−/−</sup> MEFs and the PRMT inhibitor. M.-E.T.-P. is an EMBO long-term fellow. We are grateful to the Wellcome Trust Senior Fellowship to M.Z.-G., which supported this work.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.Z.-G. (mzg@mole.bio.cam.ac.uk).



# The nuclear receptor LXR is a glucose sensor

Nico Mitro<sup>1,2</sup>, Puiying A. Mak<sup>1</sup>, Leo Vargas<sup>1</sup>, Cristina Godio<sup>1</sup>, Eric Hampton<sup>1</sup>, Valentina Molteni<sup>1</sup>, Andreas Kreusch<sup>1</sup> & Enrique Saez<sup>1,2</sup>

The liver has a central role in glucose homeostasis, as it has the distinctive ability to produce and consume glucose<sup>1</sup>. On feeding, glucose influx triggers gene expression changes in hepatocytes to suppress endogenous glucose production and convert excess glucose into glycogen or fatty acids to be stored in adipose tissue<sup>2</sup>. This process is controlled by insulin, although debate exists as to whether insulin acts directly or indirectly on the liver<sup>3</sup>. In addition to stimulating pancreatic insulin release, glucose also regulates the activity of ChREBP, a transcription factor that modulates lipogenesis<sup>4</sup>. Here we describe another mechanism whereby glucose determines its own fate: we show that glucose binds and stimulates the transcriptional activity of the liver X receptor (LXR), a nuclear receptor that coordinates hepatic lipid metabolism. D-Glucose and D-glucose-6-phosphate are direct agonists of both LXR- $\alpha$  and LXR- $\beta$ . Glucose activates LXR at physiological concentrations expected in the liver and induces expression of LXR target genes with efficacy similar to that of oxysterols, the known LXR ligands. Cholesterol homeostasis genes that require LXR for expression are upregulated in liver and intestine of fasted mice re-fed with a glucose diet, indicating that glucose is an endogenous LXR ligand. Our results identify LXR as a transcriptional switch that integrates hepatic glucose metabolism and fatty acid synthesis.

Nuclear receptors are ligand-activated transcription factors that coordinate gene expression in response to hormonal and environmental signals. Members of the superfamily that work as heterodimers with the retinoid X receptor (RXR) serve as sensors of dietary components, orchestrating the physiological response to nutrients<sup>5</sup>. LXR- $\alpha$  and LXR- $\beta$  (also called NR1H3 and NR1H2, respectively) are RXR partners that recognize oxidized cholesterol and control gene expression linked to cholesterol and fatty acid metabolism<sup>6–8</sup>. Activation of LXRs results in decreased atherosclerosis in rodents<sup>9</sup>. LXR ligands have anti-diabetic effects as well, decreasing liver glucose output and increasing peripheral glucose disposal<sup>10,11</sup>.

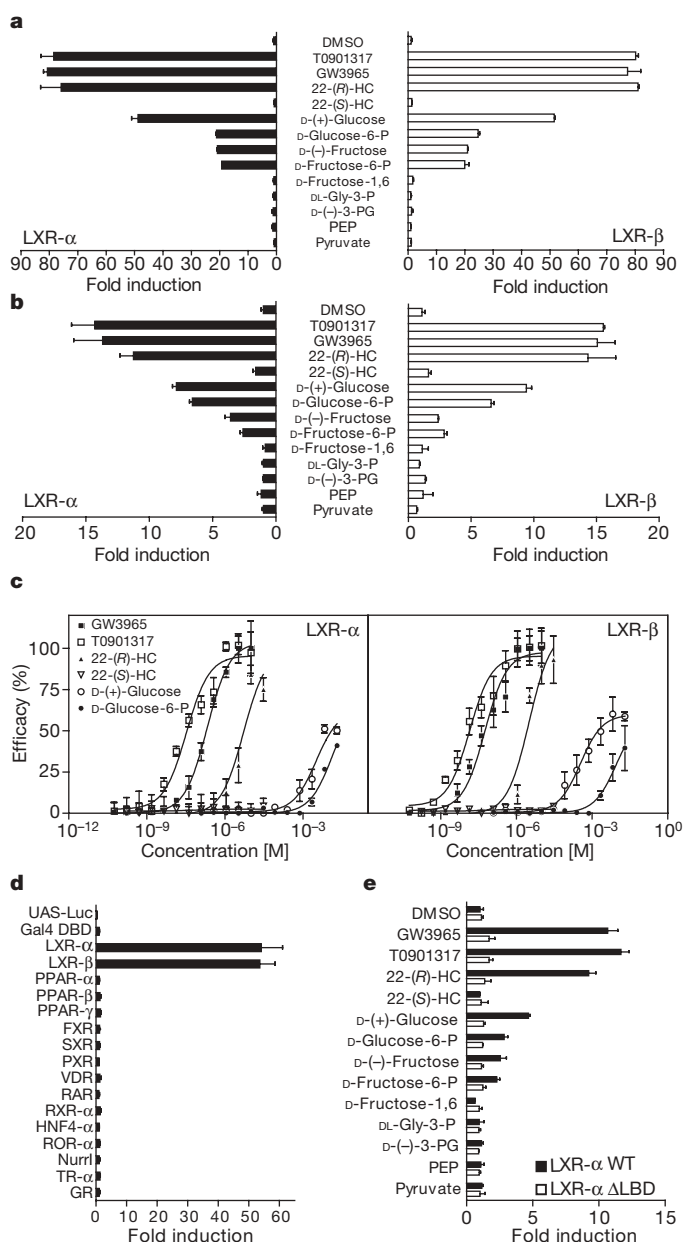
Treatment of rodents with synthetic LXR ligands results in decreased hepatic gluconeogenesis and increased lipogenesis, indicating that LXR serves as a transcription factor that integrates liver carbohydrate and lipid metabolism. Because the rodent diet is virtually devoid of cholesterol, to explore the dietary influences that modulate hepatic LXR function we asked whether LXR activity could be regulated by glucose. Human HepG2 cells were transfected with a Gal4-responsive luciferase reporter and chimaeric constructs of the ligand-binding domain (LBD) of the LXRs fused to the DNA-binding domain (DBD) of yeast Gal4. Transfected cells were maintained in media with various concentrations of glucose and subsequently exposed to increasing amounts of glucose and glycolysis derivatives. Notably, D-glucose elicited a robust 50-fold induction of LXR- $\alpha$  and LXR- $\beta$  transactivation activity when cells were grown in no glucose media (Fig. 1a). The ability of the LXRs to respond to glucose and its derivatives is very specific: no effect was seen in other nuclear receptors tested (Fig. 1d).

To examine whether glucose can regulate the LXR–RXR heterodimers, HepG2 cells were transfected with expression vectors for the LXRs, RXR and a reporter under the control of two copies of the LXR response element (LXRE) of the *ABCA1* gene. Glucose and its derivatives stimulated significant LXR–RXR activity in cells grown in no glucose (Fig. 1b). D-Glucose and D-glucose-6-phosphate are more potent on LXR- $\beta$  than LXR- $\alpha$  (effector concentration for half-maximum response ( $EC_{50}$ ) of 308  $\mu$ M versus 3,141  $\mu$ M), and are weaker inducers than known LXR ligands (Fig. 1c and Supplementary Table 1). Activation of the LXR–RXR heterodimer by glucose was dependent on the LXR LBD (Fig. 1e). Glucose concentration did not affect expression of transfected LXR protein (Supplementary Fig. 2).

Glucose and its derivatives might modulate LXR activity directly, by binding the LXR LBD, or indirectly, via the generation of an endogenous ligand or a post-transcriptional modification. To distinguish these possibilities, we tested the ability of D-glucose and glycolytic intermediates to bind LXR in a cell-free coactivator recruitment assay. In this setting, interaction between the LXR LBD and a peptide from the coactivator SRC-1, measured by fluorescence resonance energy transfer (FRET), provides an indication of ligand binding and receptor activation. Unexpectedly, D-glucose and D-glucose-6-phosphate were both able to induce coactivator recruitment, revealing that they are direct LXR agonists (Fig. 2a). D-Glucose activated the LBD of both LXRs with the same efficacy as 22-(R)-hydroxycholesterol, GW3965 and T0901317, the known LXR ligands. D-Glucose was more potent on LXR- $\beta$  than on LXR- $\alpha$  ( $EC_{50}$  of 318  $\mu$ M versus 2,904  $\mu$ M; Supplementary Fig. 3 and Supplementary Table 2). A survey of more than 40 monosaccharides using the FRET assay outlined the specificity of the effect: only D-glucose, L-glucose and D-glucose-6-phosphate brought about coactivator recruitment (Fig. 2b and data not shown).

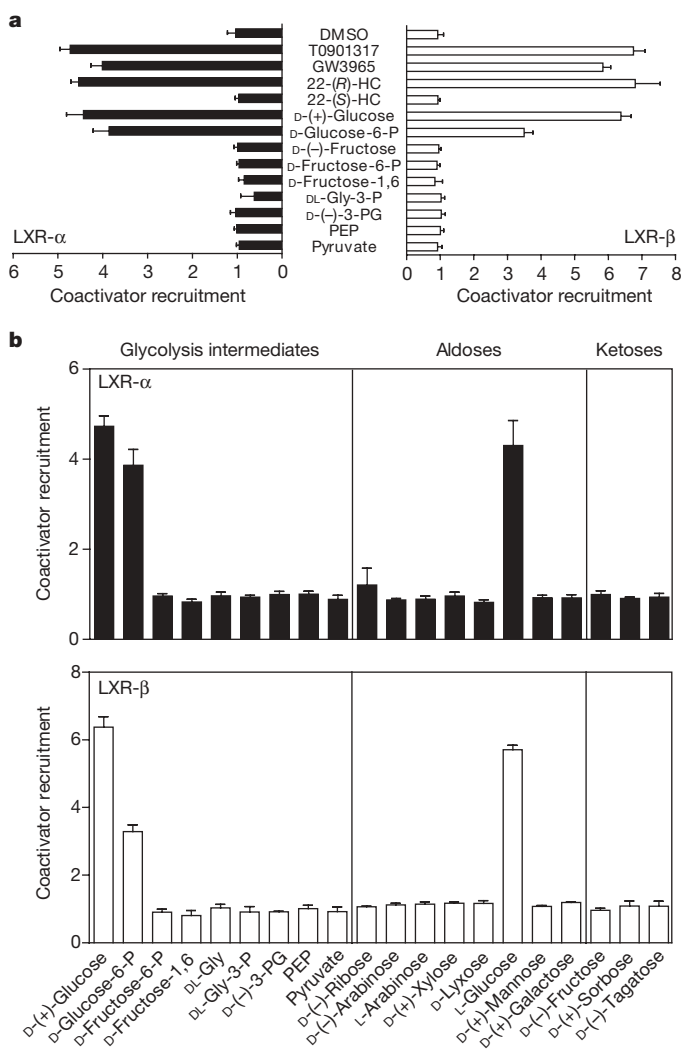
To confirm that D-glucose and D-glucose-6-phosphate bind LXR directly, a scintillation proximity assay (SPA) was used to evaluate their ability to displace a potent, labelled synthetic ligand bound to the LBD of the LXRs ( $[^3H]$ T0901317; dissociation constant ( $K_d$ ) of 7 nM for LXR- $\alpha$  and 22 nM for LXR- $\beta$ ). D-Glucose and D-glucose-6-phosphate competed effectively for binding, displacing the labelled synthetic agonist with equivalent efficacy to known ligands (D-glucose inhibition constant ( $K_i$ ) of 1,801  $\mu$ M for LXR- $\alpha$  and 202  $\mu$ M for LXR- $\beta$ ; Fig. 3a and Supplementary Table 3). To demonstrate further that D-glucose has the ability to bind directly the LXRs, the SPA assay was run using  $[^3H]$ D-glucose as the ligand. Labelled glucose bound the LBD of the LXRs and induced scintillation with a  $K_d$  of 2,489  $\mu$ M for LXR- $\alpha$  and 211  $\mu$ M for LXR- $\beta$  (Fig. 3b). Cold D-glucose and D-glucose-6-phosphate competed for binding and completely displaced  $[^3H]$ D-glucose from the LBD of the LXRs. Interestingly, the known ligands were only partially able to displace labelled glucose (Fig. 3b), suggesting that glucose can bind LXR in more than one site (see scatchard analysis, inset Fig. 3b), and that glucose and known ligands can bind LXR simultaneously. To verify this finding, an SPA

<sup>1</sup>Genomics Institute of the Novartis Research Foundation, 10675 John Hopkins Drive, San Diego, California 92121, USA. <sup>2</sup>The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, USA.



**Figure 1 | Glucose induces LXR transcriptional activity.** **a**, Activation of LXR-α and LXR-β by glucose and glycolytic intermediates. HepG2 cells transfected with a Gal4-responsive luciferase reporter and expression plasmids encoding Gal4 DBD–LXR-α/β LBD chimaeric proteins were plated in no glucose media for 24 h and treated as indicated. D-Fructose-1,6, D-fructose-1,6-bisphosphate; D-fructose-6-P, D-fructose-6-phosphate; D-glucose-6-P, D-glucose-6-phosphate; DL-Gly-3-P, DL-glyceraldehyde-3-phosphate; 22-(R)-HC, 22-(R)-hydroxycholesterol; 22-(S)-HC, 22-(S)-hydroxycholesterol; PEP, phosphoenolpyruvate; D-(-)-3-PG, D-(-)-3-phosphoglycerate. **b**, Activation profile of natural LXR–RXR heterodimers with glycolysis intermediates (20 mM top dose). Cells were transfected with a  $\times 2$  LXRE ABCA1-luciferase reporter and expression plasmids for RXR-α, and LXR-α or LXR-β. No effect was seen with pGL3 or TK-luc reporters lacking LXREs. **c**, Dose response curves on LXR-α–RXR and LXR-β–RXR of glucose and glycolytic compounds. Efficacy is relative to GW3965. **d**, Activation by glucose is LXR specific. HepG2 cells transfected with Gal4 chimaeric constructs for various nuclear receptors were treated with 20 mM glucose for 24 h. **e**, Activation by glucose requires the LXR LBD. Normalized luciferase values are expressed as fold induction versus DMSO; all error bars indicate s.d. Data are representative of four experiments performed in duplicate.

assay was run with a combination of [ $^3$ H]T0901317 and [ $^3$ H]D-glucose. The presence of labelled glucose increased scintillation well beyond the level elicited by a maximal dose of labelled T0901317, indicating that these molecules bind LXR concurrently (Fig. 3c). To explore the functional consequences of glucose binding to LXR in the presence of a known ligand, the FRET assay was run with a combination of T0901317 and glucose. Addition of glucose to a maximal dose of a known ligand resulted in increased coactivator recruitment (Fig. 3d). In contrast, addition of GW3965 to a saturating dose of T0901317 (10  $\mu$ M) did not enhance coactivator recruitment. Glucose was also able to protect LXR-α from protease digestion (Supplementary Fig. 4), and to increase the melting temperature ( $T_m$ ) of LXR-β in differential scanning calorimetry (Supplementary Fig. 5). These observations in five different biochemical assays demonstrate that D-glucose and D-glucose-6-phosphate are direct agonists of the LXRs that bind more than one site, and can work in combination with a known ligand. The precise binding mode of glucose awaits the resolution of a crystal structure.



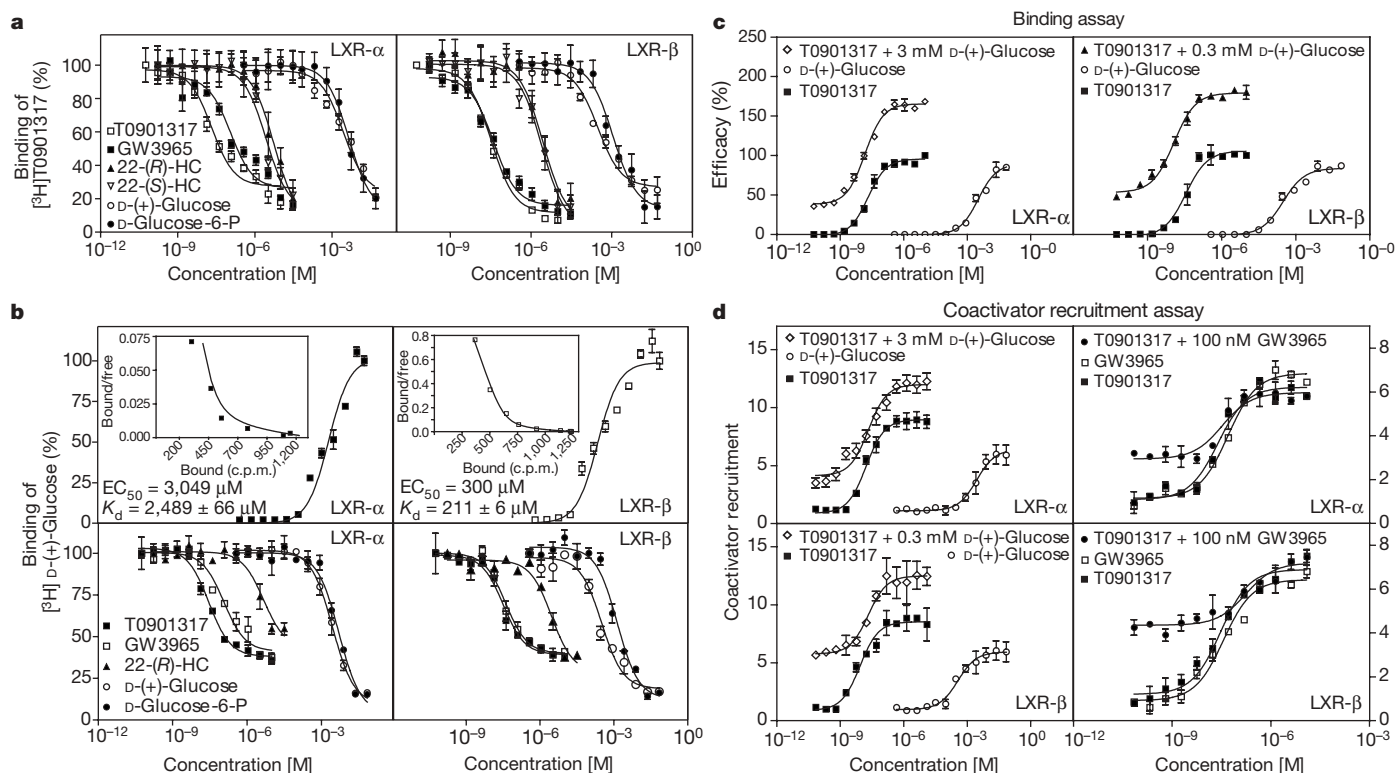
**Figure 2 | Glucose is a direct LXR agonist.** **a**, D-Glucose and glucose-6-phosphate induce SRC-1 recruitment to the LXR-α and LXR-β LBDs in cell-free FRET assays. Glycolytic intermediates were tested at 20 mM. **b**, Structure–activity relationship analysis of various monosaccharides (20 mM) on the FRET assay. Values are presented as fold induction versus vehicle (increase in coactivator recruitment measured as a change in 665/615 nm emission relative to vehicle). All error bars indicate s.d.; abbreviations as in Fig. 1. Data are representative of four experiments performed in triplicate.

To evaluate the physiological relevance of glucose as an LXR agonist, we compared the pattern of gene expression induced in HepG2 cells by D-glucose and GW3965. In cells grown in the absence of glucose or in low glucose conditions, overnight treatment with either compound (1  $\mu$ M GW3965 or 20 mM D-glucose) stimulated expression of genes involved in fatty acid synthesis and repressed expression of gluconeogenic genes (Fig. 4a). Glucose and GW3965 inhibited these genes under conditions of maximal induction (cAMP, dexamethasone) to a degree comparable to that seen with insulin, which seems to work with GW3965 and D-glucose to suppress expression of gluconeogenic genes (Supplementary Fig. 6). Interestingly, D-glucose was also able to significantly induce expression of LXR target genes involved in cholesterol homeostasis (*ABCA1*, *ABCG5*, *ABCG8*, *ABCG1*, *CETP*) that are not insulin regulated and whose expression was heretofore not associated with glucose levels (Fig. 4a). The efficacy of known LXR ligands was potentiated with increasing glucose concentration, indicating that glucose can work together with established LXR ligands. Expression of multiple control genes was not positively affected by glucose concentration, showing that health of the cells was not affected by glucose concentration in the experimental time course used. To verify that D-glucose-mediated stimulation of LXR target genes requires LXR, the experiment was repeated in HepG2 cells transfected with a short-interfering (si)RNA pool against human LXR- $\alpha$ , the dominant subtype in these cells. Induction of LXR-dependent target genes by D-glucose and other ligands was blocked in cells transfected with siRNA against LXR- $\alpha$ , but not in unperturbed cells or cells transfected with control siRNA (Supplementary Fig. 7). D-Glucose failed to upregulate LXR target genes when cells were grown in the presence of the glucose uptake inhibitor cytochalasin B, whereas GW3965 remained active, but to a

diminished degree (Supplementary Fig. 8). These results establish that D-glucose can activate endogenous LXRs and regulate genuine LXR targets in liver cells, and that it may do so together with other LXR ligands.

To extend these observations to animals, fasted mice were challenged with GW3965 or diets where the source of carbohydrate was exclusively sucrose or D-glucose. All diets were devoid of cholesterol to minimize endogenous generation of oxysterols. D-Glucose and GW3965 induced similar changes in hepatic gene expression, triggering a pattern expected to limit hepatic glucose output and increase fatty acid synthesis (Fig. 4b). D-Glucose also strongly stimulated expression of genes that require LXR for induction (for example, *Abca1*, *Abcg1*, *Spz*). Induction of insulin-independent LXR target genes was also observed with sucrose, the glucose:fructose disaccharide present in standard chow. These findings indicate that LXR functions as a glucose sensor *in vivo* that responds to increasing liver glucose uptake. To examine the effect of insulin on glucose-stimulated, LXR-related hepatic gene expression, animals rendered insulin-deficient via streptozotocin (STZ) injection were treated in an identical manner. In the absence of an increase in insulin, glucose and GW3965 were still able to induce expression of LXR-dependent genes, repress gluconeogenesis genes, and upregulate fatty acid synthesis genes (Supplementary Fig. 9). Moreover, glucose was also able to induce upregulation of bona fide LXR target genes in the intestine of wild-type and STZ-treated mice, confirming the role of glucose as a physiological LXR ligand in another tissue that faces significant glucose influx and in which the role of insulin is not as prominent (Supplementary Fig. 10 and data not shown).

The concentration of glucose required to activate LXR is within physiological range, as blood glucose concentration in healthy



**Figure 3 | Glucose displaces a labelled high-affinity LXR ligand.**

**a**, D-Glucose and glucose-6-phosphate compete for LXR binding and displace  $[^3\text{H}]\text{T}0901317$  (25 nM) in an SPA assay. **b**,  $[^3\text{H}]\text{Glucose}$  binds LXR. A scatchard analysis is shown in the inset. Labelled glucose did not bind the RXR LBD (data not shown). Unlabelled LXR ligands displace bound  $[^3\text{H}]\text{glucose}$  (20 mM) but not completely. Values are expressed as percentage binding of labelled compound. Fractional occupancy of the receptor (see Methods) is 95% for LXR- $\alpha$  and 98% for LXR- $\beta$ . c.p.m., counts per minute.

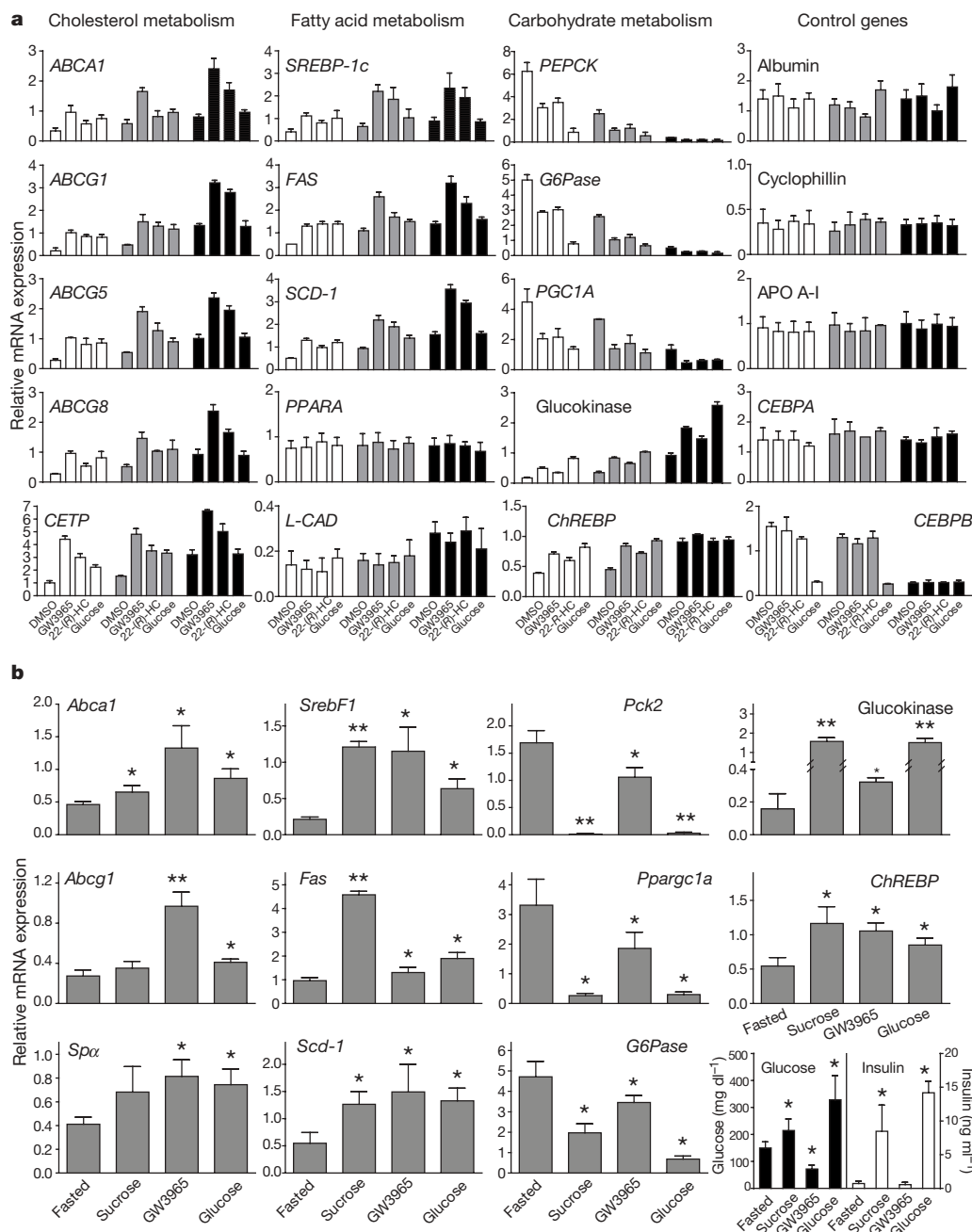
**c**, Addition of labelled glucose to a saturating dose of  $[^3\text{H}]\text{T}0901317$  (10  $\mu$ M) increases scintillation in an SPA assay; percentage efficacy is relative to 10  $\mu$ M  $[^3\text{H}]\text{T}0901317$ . **d**, Addition of glucose, but not GW3965, to a maximal dose of T0901317 enhances coactivator recruitment. Note the different scales. Values are presented as fold induction versus vehicle (increase in coactivator recruitment measured as a change in 665/615 nm emission relative to vehicle). All error bars indicate s.d.; experiments performed in triplicate.



individuals can vary from 3 to 9 mM during the day<sup>1</sup>. In tissues that express the high-capacity glucose transporter GLUT2 (liver, pancreas, kidney; Michaelis constant ( $K_m$ ) of 20–40 mM), glucose enters the cell readily and can diffuse freely to the nucleus, for the nuclear pore displays passive permeability for molecules <500 Da<sup>12</sup>. Direct measurement of nuclear glucose concentration using live imaging of FRET-based glucose nanosensors in kidney cells indicates that it is approximately 50% that of the external medium<sup>13</sup>.

Insulin promotes hepatic lipogenesis through induction of *SREBP-1c*<sup>14,15</sup>. Deletion of this SREBP isoform results in impaired fatty acid and triglyceride synthesis<sup>16</sup>. *SREBP-1c* expression is abolished in LXR- $\alpha/\beta$  null animals, showing that the LXRs are primary

transcriptional regulators of this gene<sup>17</sup>. Recent work has shown that insulin-mediated activation of *SREBP-1c* expression requires LXR, leading to the hypothesis that insulin may regulate *SREBP-1c* expression through the production of an LXR ligand<sup>18</sup>. We would like to suggest that glucose itself is the LXR ligand involved in this response (Supplementary Fig. 1). LXR could also influence hepatic glucose fate via its ability to induce expression of ChREBP (Fig. 4a, b), although activity of this transcription factor is regulated at the post-transcriptional level. Deletion of ChREBP reduces but does not ablate lipogenesis<sup>19</sup>, indicating that concerted action of *SREBP-1c* and ChREBP is necessary for normal fatty acid and triglyceride synthesis *in vivo*.



**Figure 4 | Glucose regulates direct LXR target genes *in vivo*.** **a**, HepG2 cells cultured in 0 mM (white bars), 2 mM (grey bars) or 25 mM (black bars) glucose medium were treated overnight with GW3965 (1 μM), 22-(R)-hydroxycholesterol (5 μM), or D-glucose (20 mM) and gene expression was analysed using qRT-PCR. Glucose stimulates expression of direct LXR cholesterol homeostasis target genes. Note that efficacy of known LXR ligands increases with increasing glucose concentration. **b**, Glucose induces

LXR target genes in mouse liver. Mice fasted overnight were challenged orally with GW3965 (50 mg kg<sup>-1</sup>), or re-fed with a glucose or sucrose diet and killed 6 h later. D-Glucose and GW3965 regulate the same direct LXR targets (genes involved in cholesterol and fatty acid metabolism) as well as indirect carbohydrate metabolism targets. All error bars represent s.d.,  $n = 5-6$  mice per group. Asterisk,  $P < 0.05$ ; double asterisk,  $P < 0.001$  treatment versus fasted.

The integration of glucose sensing and control of lipogenesis in the same protein may provide an explanation for the observation that low-fat, high-carbohydrate diets induce hypertriglyceridaemia<sup>20</sup>: LXR can sense surplus glucose, induce fatty acid synthesis, and prompt hepatic export of very low density lipoprotein (VLDL)<sup>21</sup>. Because LXR acts as the body's sensor of pathogenic cholesterol build-up, its ability to bind both glucose and oxysterols also suggests that LXR may be the link between hyperglycaemia and atherosclerosis. However, it remains to be established whether enough intracellular D-glucose or D-glucose-6-phosphate is present to activate LXR in tissues other than liver and intestine.

## METHODS

Detailed protocols for all assays can be found in Supplementary Information.

Cell culture and transcriptional assays were performed essentially as described<sup>22</sup>. Binding assays (FRET and SPA) were performed as described in Supplementary Information and ref. 23. Binding data were analysed as in ref. 24. Animal experiments (described in Supplementary Information) were approved by the Institutional Animal Care and Use Committee of The Genomics Institute of the Novartis Research Foundation. Gene expression was analysed using Taqman qRT-PCR (see Supplementary Information).

Received 25 July; accepted 7 November 2006.

Published online 24 December 2006.

- Zierler, K. Whole body glucose metabolism. *Am. J. Physiol.* **276**, E409–E426 (1999).
- Moore, M. C., Cherrington, A. D. & Wasserman, D. H. Regulation of hepatic and peripheral glucose disposal. *Best Pract. Res. Clin. Endocrinol. Metab.* **17**, 343–364 (2003).
- Cherrington, A. D. The role of hepatic insulin receptors in the regulation of glucose production. *J. Clin. Invest.* **115**, 1136–1139 (2005).
- Yamashita, H. *et al.* A glucose-responsive transcription factor that regulates carbohydrate metabolism in the liver. *Proc. Natl Acad. Sci. USA* **98**, 9116–9121 (2001).
- Shulman, A. I. & Mangelsdorf, D. J. Retinoid x receptor heterodimers in the metabolic syndrome. *N. Engl. J. Med.* **353**, 604–615 (2005).
- Schultz, J. R. *et al.* Role of LXRs in control of lipogenesis. *Genes Dev.* **14**, 2831–2838 (2000).
- Peet, D. J. *et al.* Cholesterol and bile acid metabolism are impaired in mice lacking the nuclear oxysterol receptor LXR $\alpha$ . *Cell* **93**, 693–704 (1998).
- Janowski, B. A., Willy, P. J., Devi, T. R., Falck, J. R. & Mangelsdorf, D. J. An oxysterol signalling pathway mediated by the nuclear receptor LXR $\alpha$ . *Nature* **383**, 728–731 (1996).
- Joseph, S. B. *et al.* Synthetic LXR ligand inhibits the development of atherosclerosis in mice. *Proc. Natl Acad. Sci. USA* **99**, 7604–7609 (2002).
- Laffitte, B. A. *et al.* Activation of liver X receptor improves glucose tolerance through coordinate regulation of glucose metabolism in liver and adipose tissue. *Proc. Natl Acad. Sci. USA* **100**, 5419–5424 (2003).
- Cao, G. *et al.* Antidiabetic action of a liver x receptor agonist mediated by inhibition of hepatic gluconeogenesis. *J. Biol. Chem.* **278**, 1131–1136 (2003).
- Stehno-Bittel, L., Perez-Terzic, C. & Clapham, D. E. Diffusion across the nuclear envelope inhibited by depletion of the nuclear Ca<sup>2+</sup> store. *Science* **270**, 1835–1838 (1995).
- Fehr, M., Lalonde, S., Ehrhardt, D. W. & Frommer, W. B. Live imaging of glucose homeostasis in nuclei of COS-7 cells. *J. Fluoresc.* **14**, 603–609 (2004).
- Shimomura, I. *et al.* Insulin selectively increases SREBP-1c mRNA in the livers of rats with streptozotocin-induced diabetes. *Proc. Natl Acad. Sci. USA* **96**, 13656–13661 (1999).
- Eberle, D., Hegarty, B., Bossard, P., Ferre, P. & Foulle, F. SREBP transcription factors: master regulators of lipid homeostasis. *Biochimie* **86**, 839–848 (2004).
- Liang, G. *et al.* Diminished hepatic response to fasting/refeeding and liver X receptor agonists in mice with selective deficiency of sterol regulatory element-binding protein-1c. *J. Biol. Chem.* **277**, 9520–9528 (2002).
- Repa, J. J. *et al.* Regulation of mouse sterol regulatory element-binding protein-1c gene (SREBP-1c) by oxysterol receptors, LXR $\alpha$  and LXR $\beta$ . *Genes Dev.* **14**, 2819–2830 (2000).
- Chen, G., Liang, G., Ou, J., Goldstein, J. L. & Brown, M. S. Central role for liver X receptor in insulin-mediated activation of Srebp-1c transcription and stimulation of fatty acid synthesis in liver. *Proc. Natl Acad. Sci. USA* **101**, 11245–11250 (2004).
- Iizuka, K., Bruick, R. K., Liang, G., Horton, J. D. & Uyeda, K. Deficiency of carbohydrate response element-binding protein (ChREBP) reduces lipogenesis as well as glycolysis. *Proc. Natl Acad. Sci. USA* **101**, 7281–7286 (2004).
- Parks, E. J. Changes in fat synthesis influenced by dietary macronutrient content. *Proc. Nutr. Soc.* **61**, 281–286 (2002).
- Lin, J. *et al.* Hyperlipidemic effects of dietary saturated fats mediated through PGC-1 $\beta$  coactivation of SREBP. *Cell* **120**, 261–273 (2005).
- Joseph, S. B. *et al.* LXR-dependent gene expression is important for macrophage survival and the innate immune response. *Cell* **119**, 299–309 (2004).
- Janowski, B. A. *et al.* Structural requirements of ligands for the oxysterol liver X receptors LXR $\alpha$  and LXR $\beta$ . *Proc. Natl Acad. Sci. USA* **96**, 266–271 (1999).
- DeBlasi, A., O'Reilly, K. & Motulsky, H. J. Calculating receptor number from binding experiments using same compound as radioligand and competitor. *Trends Pharmacol. Sci.* **10**, 227–229 (1989).

Supplementary Information is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank S. Bohan, R. Romeo, B. Geierstanger, M. Chalmers, P. Griffin, N. Gekakis, M. Crestani, H. Shimano, T. Matsuzaka, P. Tontonoz, B. Laffitte, S. Joseph, A. Brock, E. Peters and K. Nettles for technical help and/or useful comments. C.G. was a visiting scientist supported by a fellowship from the Department of Pharmacological Sciences, University of Milano, Italy.

**Author Contributions** L.V., C.G., E.H. and A.K. performed experiments; P.A.M. and V.M. designed and performed experiments and analysed data; and N.M. and E.S. designed and performed experiments, analysed data, and wrote the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to E.S. ([esaez@scripps.edu](mailto:esaez@scripps.edu)).

# naturejobs

**THE CAREERS  
MAGAZINE FOR  
SCIENTISTS**

**N**ot so long ago, contract-research organizations (CROs) were content to be simple fee-for-service businesses. They would test drugs in animals, run clinical trials and deal with regulatory issues. Now, as drug companies are looking to cut costs, CROs show signs of taking on more risks and responsibilities — in exchange for bigger rewards. This trend has implications not just for the financial bottom line of the drug companies, but also for scientists with aspirations to work in drug development.

One arm of Quintiles of Research Triangle Park, North Carolina, one of the world's largest CROs with 17,000 employees worldwide, is combining the approach of a traditional service provider with that of a venture-capital firm. The Quintiles subsidiary NovaQuest will now take on the financial costs of late-stage development, testing and getting approval for a pharmaceutical partner's drug candidates, with the understanding that NovaQuest will get a return from products that reach the market.

This approach is appealing to drugs companies, as it allows them to reduce their workforce and focus on drug discovery and early-stage development, while letting the CRO share the risk on the clinical trials, regulatory affairs and marketing. But it is worrying for people working in or aspiring to work in the pharmaceutical industry — as it reinforces earlier outsourcing trends where companies have closed down Western facilities and moved some activities to the developing world.

Eli Lilly and Pfizer have already signed up to the new co-promotion approach, says Ron Wooten, NovaQuest's president. As more clients come on board, the number of positions at drug companies in areas such as clinical-trial management and regulatory affairs is likely to shrink or, as Wooten puts it, "reduce permanent headcount". Meanwhile, more of these positions should be opening up at CROs — especially if others adopt Quintiles' strategy. The question remains, however, whether there will be just a sideways shift or an eventual net loss of late-stage drug-development jobs as outsourcing and consolidation continue.

**Paul Smaglik, *Naturejobs* editor**

**Editor:** Paul Smaglik  
**Assistant Editor:** Gene Russo

**European Head Office, London**  
The Macmillan Building,  
4 Crinan Street,  
London N1 9XW, UK  
Tel: +44 (0) 20 7843 4961  
Fax: +44 (0) 20 7843 4996  
e-mail: [naturejobs@nature.com](mailto:naturejobs@nature.com)

**European Sales Manager:**  
Andy Douglas (4975)  
e-mail: [a.douglas@nature.com](mailto:a.douglas@nature.com)  
**Business Development Manager:**  
Amelie Pequignot (4974)  
e-mail: [a.pequignot@nature.com](mailto:a.pequignot@nature.com)  
**Natureevents:**  
Claudia Paulsen Young (+44 (0) 20 7014 4015)  
e-mail: [c.paulsenyoung@nature.com](mailto:c.paulsenyoung@nature.com)  
**France/Switzerland/Belgium:**  
Muriel Lestringuez (4994)

**UK/Ireland/Italy/RoW:**  
Loredana Milanese (4944)  
Nils Moeller (4953)  
**Scandinavia/Spain/Portugal:**  
Evelina Rubio-Morgan (4973)  
**Germany/Austria/The Netherlands:**  
Reya Silao (4970)  
**Online Job Postings:**  
Matthew Ward (+44 (0) 20 7014 4059)

**European Satellite Office**  
**Germany:** Patrick Phelan  
Tel: +49 89 54 90 57 11  
Fax: +49 89 54 90 57 20  
e-mail: [p.phelan@nature.com](mailto:p.phelan@nature.com)

**Advertising Production Manager:**  
Stephen Russell  
To send materials use London address above.  
Tel: +44 (0) 20 7843 4816  
Fax: +44 (0) 20 7843 4996  
e-mail: [naturejobs@nature.com](mailto:naturejobs@nature.com)

**Naturejobs web development:** Tom Hancock  
**Naturejobs online production:**  
Catherine Alexander

**US Head Office, New York**  
75 Varick Street, 9th Floor,  
New York, NY 10013-1917  
Tel: +1 800 989 7718  
Fax: +1 800 989 7103  
e-mail: [naturejobs@natureny.com](mailto:naturejobs@natureny.com)

**US Sales Manager:** Peter Bless

**Japan Head Office, Tokyo**  
Chiyoda Building, 2-37 Ichigayatamachi,  
Shinjuku-ku, Tokyo 162-0843  
Tel: +81 3 3267 8751  
Fax: +81 3 3267 8746

**Asia-Pacific Sales Manager:**  
Ayako Watanabe  
e-mail: [a.watanabe@natureasia.com](mailto:a.watanabe@natureasia.com)



# Following the law

A guide for the perplexed graduate student doing research.



Irving P. Herman

Going to graduate school to pursue a doctorate is a major commitment of time and effort. It is not for everyone. Once in a graduate programme, choosing a research adviser is perhaps the most important decision a student can make. Likewise, choosing the right students is essential for the careers of advisers. It is obvious that mentors and students must have common research interests and compatible work habits. But it is sometimes less obvious that they must learn to communicate with each other. Developing a functional working relationship is important even between 'good' students and 'good' advisers, and this often takes some time.

In this spirit, I offer 20 'laws' as a guide to graduate students doing thesis research. Each contains sound advice about the facts of life in graduate research, particularly from the viewpoint of a thesis adviser. Several have been slightly exaggerated for effect, or are not to be taken too literally. Some clearly pertain to experimental research, although they have obvious counterparts for other types of research.

I developed these laws to help motivate some of the graduate students in my group, to explain how to be an effective student and to convince them that supervised research is a symbiotic (although not symmetric) interaction between student and adviser. I admit that I am not always successful in this endeavour.

I also use these laws as general advice for graduate students in

**"Supervised research is a symbiotic (although not symmetric) interaction between student and adviser."**

my department. All doctoral candidates in the department receive a copy of the laws when they enter our programme, to help them understand how to work with an adviser as they move from the undergraduate mode of taking courses to the graduate mode of conducting supervised research. These laws seem to resonate with my faculty colleagues, and several have posted them and used them.

Although the laws require

no interpretation, students may understand them better by understanding their advisers better. Advisers, including my own students' adviser, love to recall the 'good old days' when they were graduate students. They all worked 20-hour days, seven days a week, and they never slept. They needed to build from scratch every instrument they used in their work and they thought of every idea in their theses. And, most importantly, they always, always, took to heart their advisers' every suggestion and acted on each promptly.

On a more serious note, there are some real overarching themes in the laws. In research, being right is paramount, and ideas and results must be evaluated using objective methods untainted by egos. Productive people are productive because they have good work habits. Students need to grow professionally and advisers need to assist them. The relationship between graduate student and adviser truly is symbiotic.

**Irving P. Herman is a professor of applied physics at Columbia University, New York.**

## THE LAWS OF HERMAN

1. Your vacation begins after you defend your thesis.
2. In research, what matters is what is right, and not who is right.
3. In research and other matters, your adviser is always right, most of the time.
4. Act as if your adviser is always right, almost all the time.
5. If you think you are right and you are able to convince your adviser, your adviser will be very happy.
6. Your productivity varies as (effective productive time spent per day)<sup>1,000</sup>.
7. Your productivity also varies as 1/(your delay in analysing acquired data)<sup>1,000</sup>.
8. Take data today as if you know that your equipment will break tomorrow.
9. If you would be unhappy to lose your data, make a permanent back-up copy of them within five minutes of acquiring them.
10. Your adviser expects your productivity to be low initially and then to be above threshold after a year or so.
11. You must become a bigger expert in your thesis area than your adviser.
12. When you cooperate, your adviser's blood pressure will go down a bit.
13. When you don't cooperate, your adviser's blood pressure either goes up a bit or it goes down to zero.
14. Usually, only when you can publish your results are they good enough to be part of your thesis.
15. The higher the quality, first, and quantity, second, of your publishable work, the better your thesis.
16. Remember, it's your thesis. You (!) need to do it.
17. Your adviser wants you to become famous, so that he/she can finally become famous.
18. Your adviser wants to write the best letter of recommendation for you that is possible.
19. Whatever is best for you is best for your adviser.
20. Whatever is best for your adviser is best for you.

These laws were inspired by the 'Laws of the House of God' from *The House of God* by Samuel Shem (Richard Marek, 1978), which provided a somewhat different brand of advice to medical interns. The author thanks Jonathan Spanier, Yigal Komem and other colleagues for suggestions.